

B4

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
1 November 2001 (01.11.2001)

PCT

(10) International Publication Number  
**WO 01/81555 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 9/00**

(21) International Application Number: **PCT/US01/12992**

(22) International Filing Date: 20 April 2001 (20.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/199,021	20 April 2000 (20.04.2000)	US
60/200,226	28 April 2000 (28.04.2000)	US
60/202,339	5 May 2000 (05.05.2000)	US
60/203,505	11 May 2000 (11.05.2000)	US
60/205,564	18 May 2000 (18.05.2000)	US
60/207,739	26 May 2000 (26.05.2000)	US
60/208,795	1 June 2000 (01.06.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/200,226 (CIP)
Filed on	28 April 2000 (28.04.2000)
US	60/199,021 (CIP)
Filed on	20 April 2000 (20.04.2000)
US	60/202,339 (CIP)
Filed on	5 May 2000 (05.05.2000)
US	60/203,505 (CIP)
Filed on	11 May 2000 (11.05.2000)
US	60/205,564 (CIP)
Filed on	18 May 2000 (18.05.2000)
US	60/207,739 (CIP)
Filed on	26 May 2000 (26.05.2000)
US	60/208,795 (CIP)
Filed on	1 June 2000 (01.06.2000)

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 940304 (US)**

(72) Inventors; and

(75) Inventors/Applicants (for US only): **YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). GRIFFIN, Jennifer,**

A. [US/US]; 181 Irene Court #11, Belmont, CA 94002 (US). NGUYEN, Dannielle, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). KHAN, Farrah, A. [IN/US]; 333 Escuela Avenue, #221, Mountain View, CA 94040 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). BURRILL, John, D. [US/US]; 650 Emerald Hill Road, Redwood City, CA 94061 (US). MARCUS, Gregory, A. [US/US]; 1714 Connecticut Drive, Redwood City, CA 94061 (US). ZINGLER, Kurt, A. [US/US]; 723 Ashbury Street #C, San Francisco, CA 94117 (US). RECIPON, Shirley, A. [US/US]; 85 Fortuna Avenue, San Francisco, CA 95115-3873 (US). LU, Yan [CN/US]; 3885 Cormia Way, Palo Alto, CA 94303 (US). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). HAFALIA, April [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polten Place Way, San Jose, CA 95121 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury Kent CT1 3EZ (GB). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US). GURURAJAN, Rajagopal [US/US]; 5591 Dent Avenue, San Jose, CA 95118 (US).

(74) Agents: **HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US)**

(81) Designated States (national): **AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,**

*[Continued on next page]*

(54) Title: **HUMAN KINASES**

(57) Abstract: The invention provides human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.

**WO 01/81555 A2**



LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

**Published:**

— without international search report and to be republished  
upon receipt of that report

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## HUMAN KINASES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human kinases and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

### 10 BACKGROUND OF THE INVENTION

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

30 There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing 35 specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic

domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains

5 specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

10 In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) *The Protein Kinase Facts Book*, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the 15 kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

#### Protein Tyrosine Kinases

20 Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, 25 platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and 30 prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493). Regulation of PTK activity may therefore be an important strategy in

controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism, and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a

highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) *J. Biol. Chem.* 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, 5  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . Fish et al., identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

10 The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.

15 Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al., have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian

20 rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) *Science* 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al., (2000) *Biochimie* 82:1123-7). HIPKs contain a conserved protein kinase domain separated from a domain that interacts with 25 homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al., (1998) *J. Biol. Chem.* 273:25875-9; Wang, Y. et al., (2001) *Biochim. Biophys. Acta* 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each are important in the regulation of cellular protein function (Kim, Y.H. et al., (1999) *Proc. Nat. Acad. Sci. U.S.A.* 96:12350-5).

#### Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium 35 receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also

activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

#### Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). MAP kinase signaling pathways are present in mammalian cells as well as in yeast. The extracellular stimuli which activate MAP kinase pathways include epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.

#### Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation 35 of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO

J. 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) *Science* 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) *Science* 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A similar deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) *J. Biol. Chem.* 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) *J. Biol. Chem.* 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune disease, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a

5 C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) *J. Biol. Chem.* 273:29066-29071). DRAK1 and DRAK2 are STKs that share

10 homology with the death-associated protein kinases (DAP kinases), known to function in interferon- $\gamma$  induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the

15 N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) *J. Biol. Chem.* 273:12296-12300). CD95 is a

20 member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) *Cell* 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal

25 kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

30 **Mitochondrial Protein Kinases**

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-

35 dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for

phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) *Adv. Enzyme Regul.* 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory 5 enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) *Adv. Enzyme Regul.* 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of 10 the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

#### KINASES WITH NON-PROTEIN SUBSTRATES

15 Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) *Curr. Opin. Cell. Biol.* 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the 20 protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP<sub>2</sub>). PIP<sub>2</sub> is then cleaved into inositol triphosphate (IP<sub>3</sub>) and diacylglycerol. These two 25 products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. 30 PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP<sub>2</sub>) to PI (3,4,5) P<sub>3</sub> (PIP<sub>3</sub>). PIP<sub>3</sub> then activates a number of other proteins, 35 including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3,

and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) *Biochem. J.* 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role 5 for PI3K in type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) *J. Biol. Chem.* 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein 10 coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including 15 platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., *supra*).

#### Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and 20 guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative 25 drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) *Cancer Res.* 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated 30 with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) *J. Biol. Chem.* 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity to treat certain cancers. Alternatively, reduced AdK activity may be a source of various 35 metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be

treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21<sup>ras</sup> known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) *Cancer Res.* 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21<sup>ras</sup> and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and R.L. Miller (1980) *J. Biol. Chem.* 255:7204-7207; Stenberg, K. et al. (1986) *J. Biol. Chem.* 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

#### Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) *Pharmacol. Ther.* 67:155-186).

The discovery of new human kinases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

#### **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, human kinases, referred to collectively as "PKIN" and individually as "PKIN-1," "PKIN-2," "PKIN-3," "PKIN-4," "PKIN-5," "PKIN-6," "PKIN-7," "PKIN-8," "PKIN-9," "PKIN-10," "PKIN-11," "PKIN-12," "PKIN-13," "PKIN-14,"

“PKIN-15,” “PKIN-16,” “PKIN-17,” and “PKIN-18.” In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-18.

5 The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-18. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:19-36.

10 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

15 The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino

20

25

30

35

acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

5        Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an 10      amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

15      The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

20      Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of 25      SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a 30      hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

35      The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting

of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide 5 complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a 10 polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic 15 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the 20 composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence 25 selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the 30 invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as

an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a

5 polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a

10 pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ

15 ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the

20 group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of

the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:19-36, the method comprising a) 5 exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 10 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the 15 polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 20 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target 25 polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

#### BRIEF DESCRIPTION OF THE TABLES

30 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

5 Table 4 lists the cDNA and genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

10 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

15 Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

20 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

25 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing 30 the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

35 "PKIN" refers to the amino acid sequences of substantially purified PKIN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and

human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PKIN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with 5 PKIN or by acting on components of the biological pathway in which PKIN participates.

An "allelic variant" is an alternative form of the gene encoding PKIN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in 10 polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to 15 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PKIN include those sequences with deletions, 15 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PKIN or a polypeptide with at least one functional characteristic of PKIN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PKIN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding 20 PKIN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PKIN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PKIN is retained. For example, 25 negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, 30 polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. 35 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well

known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PKIN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by 5 directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PKIN polypeptides can be prepared using intact polypeptides or using fragments 10 containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

15 The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen 20 used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified 25 sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or 30 translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PKIN, or of any oligopeptide thereof, 35 to induce a specific immune response in appropriate animals or cells and to bind with specific

antibodies.

“Complementary” describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

5 A “composition comprising a given polynucleotide sequence” and a “composition comprising a given amino acid sequence” refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PKIN or fragments of PKIN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be

10 associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

“Consensus sequence” refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied

15 Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

20 “Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
30	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
35	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr

5	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of 10 the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an 15 alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a 20 measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

25 A "fragment" is a unique portion of PKIN or the polynucleotide encoding PKIN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 30 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the 35 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:19-36 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:19-36, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:19-36 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish

5 SEQ ID NO:19-36 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:19-36 and the region of SEQ ID NO:19-36 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-18 is encoded by a fragment of SEQ ID NO:19-36. A fragment of SEQ ID NO:1-18 comprises a region of unique amino acid sequence that specifically identifies

10 SEQ ID NO:1-18. For example, a fragment of SEQ ID NO:1-18 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-18. The precise length of a fragment of SEQ ID NO:1-18 and the region of SEQ ID NO:1-18 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

15 A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

20 The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

25 Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 30 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

35 Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment

Search Tool (BLAST) (Altschul, S.F. et al. (1990) *J. Mol. Biol.* 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with 5 other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to 10 compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*  
*Reward for match: 1*  
*Penalty for mismatch: -2*  
15 *Open Gap: 5 and Extension Gap: 2 penalties*  
*Gap x drop-off: 50*  
*Expect: 10*  
*Word Size: 11*  
*Filter: on*  
20 Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length 25 supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid 30 sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative 35 substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the

site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of 5 polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise 10 comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

15 *Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, 20 for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be 25 used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid 30 sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. 35 Specific hybridization complexes form under permissive annealing conditions and remain hybridized

after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_{ot}$  or  $R_{ot}$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

5 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PKIN which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PKIN which is useful in any of the antibody production methods disclosed herein or known in the art.

10 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

15 The term "modulate" refers to a change in the activity of PKIN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PKIN.

20 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

25 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

30 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

35 "Post-translational modification" of an PKIN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PKIN.

35 "Probe" refers to nucleic acid sequences encoding PKIN, their complements, or fragments

thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target 5 polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also 10 be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the, specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for 15 example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that 20 purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 25 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 30 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific 35 needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping

Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and

5 polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence  
10 that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a  
15 recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a  
vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is  
20 expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

25 "Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear  
30 sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PKIN, nucleic acids encoding PKIN, or fragments thereof may comprise a bodily fluid; an extract from a  
35 cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or

cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in

vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, 5 transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of 10 the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an 15 "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to 20 another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a 25 propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at 30 least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

35 The invention is based on the discovery of new human kinases (PKIN), the polynucleotides

encoding PKIN, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are human kinases. For example, SEQ ID NO:12 is 95% identical to a human adenylate kinase (GenBank ID g28577) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2). The BLAST probability score is 2.4e-112, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLIMPS, MOTIFS, and PROFILE analysis; from BLAST analysis using the DOMO and PRODOM databases; and from HMMER analysis using the PFAM database further support the categorization of SEQ ID NO:12 as an adenylate kinase. (See Table 3). In an alternative example,

SEQ ID NO:13 is 92% identical to rat PCTAIRE 3 (GenBank ID g2257588) as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 1.4e-210, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance (Table 2). PCTAIRE 1, 2, and 3 comprise a subfamily of Cdc2-related kinases that are primarily expressed 5 in post-mitotic cells. PCTAIRE 2 and PCTAIRE 3 are expressed in the brain (Hirose, T. *et al.*, (1997) Eur. J. Biochem. 249:481-488; Okuda, T. *et al.*, (1992) Oncogene 7:2249-2258). SEQ ID NO:13 also contains a kinase active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative 10 evidence that SEQ ID NO:13 is a kinase. In another alternative example, SEQ ID NO:18 is 86% identical to human cell cycle related kinase (GenBank ID g4090958) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 9.2e-85, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:18 also contains a eukaryotic protein kinase domain as determined by searching for 15 statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:18 is a cell cycle related kinase. SEQ ID NO:1-11 and SEQ ID NO:14-17 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-18 are described in Table 7.

20 As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention.

25 Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:19-36 or that distinguish between SEQ ID NO:19-36 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages

30 comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

35 The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6311370H1 is the

identification number of an Incyte cDNA sequence, and NERDTDN03 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70518523D1). Alternatively, the identification numbers in 5 may refer to GenBank cDNAs or ESTs (e.g., g1860144) which contributed to the assembly of the full length 5 polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g5924006\_004.edit is the identification number of a Genscan-predicted coding sequence, with g5924006 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.)

10 Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon-stretching” algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was 15 obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which 20 were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PKIN variants. A preferred PKIN variant is one which has at least about 80%, or alternatively at least about 90%, or alternatively at least about 95%, or even at least about 98% amino acid sequence identity to the PKIN amino acid sequence, and which contains 25 at least one functional or structural characteristic of PKIN.

The invention also encompasses polynucleotides which encode PKIN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36, which encodes PKIN. The polynucleotide sequences of SEQ ID NO:19-36, as presented in the Sequence Listing, embrace the equivalent RNA sequences, 30 wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PKIN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or even at least about 98% polynucleotide 35 sequence identity to the polynucleotide sequence encoding PKIN. A particular aspect of the

invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36 which has at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or even at least about 98% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:19-36. Any one of the 5 polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PKIN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PKIN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be 10 produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PKIN, and all such variations are to be considered as being specifically disclosed.

15 Although nucleotide sequences which encode PKIN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PKIN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PKIN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide 20 occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PKIN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

25 The invention also encompasses production of DNA sequences which encode PKIN and PKIN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PKIN or any fragment thereof.

30 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:19-36 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in 35 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or

5 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA

10 sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

15 The nucleic acid sequences encoding PKIN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.)

20 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom,

25 M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries

30 (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of

35 about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence 5 into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the 10 emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

15 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PKIN may be cloned in recombinant DNA molecules that direct expression of PKIN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PKIN.

20 The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PKIN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction 25 sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or 30 improve the biological properties of PKIN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired 35 properties. These preferred variants may then be pooled and further subjected to recursive rounds of

DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of 5 homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PKIN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids 10 Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, PKIN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis 15 may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PKIN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid 20 chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active PKIN, the nucleotide sequences encoding PKIN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains 25 the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PKIN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PKIN. Such signals 30 include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PKIN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be 35 provided by the vector. Exogenous translational elements and initiation codons may be of various

origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression

5 vectors containing sequences encoding PKIN and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

10 A variety of expression vector/host systems may be utilized to contain and express sequences encoding PKIN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New

15 York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.)

20 The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PKIN. For example, routine cloning,

30 subcloning, and propagation of polynucleotide sequences encoding PKIN can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSORT1 plasmid (Life Technologies). Ligation of sequences encoding PKIN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for

35 *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of

nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of PKIN are needed, e.g. for the production of antibodies, vectors which direct high level expression of PKIN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

5 Yeast expression systems may be used for production of PKIN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 10 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of PKIN. Transcription of sequences encoding PKIN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology 20 (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PKIN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain 25 infective virus which expresses PKIN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of 30 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression 35 of PKIN in cell lines is preferred. For example, sequences encoding PKIN can be transformed into

cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; Lowy, I. et al. (1980) *Cell* 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PKIN is inserted within a marker gene sequence, transformed cells containing sequences encoding PKIN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PKIN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PKIN and that express PKIN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

35 Immunological methods for detecting and measuring the expression of PKIN using either

specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PKIN is preferred, but a

5 competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

10 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PKIN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PKIN, or any fragments thereof, may be cloned into a vector

15 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for

20 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PKIN may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence

25 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PKIN may be designed to contain signal sequences which direct secretion of PKIN through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of

30 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the

35 American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct

modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PKIN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PKIN protein

5 containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PKIN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG,

10 *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a

15 proteolytic cleavage site located between the PKIN encoding sequence and the heterologous protein sequence, so that PKIN may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

20 In a further embodiment of the invention, synthesis of radiolabeled PKIN may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

25 PKIN of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PKIN. At least one and up to a plurality of test compounds may be screened for specific binding to PKIN. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of

30 PKIN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PKIN binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for

35 these compounds involves producing appropriate cells which express PKIN, either as a secreted

protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PKIN or cell membrane fractions which contain PKIN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PKIN or the compound is analyzed.

5 An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PKIN, either in solution or affixed to a solid support, and detecting the binding of PKIN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a

10 labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PKIN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PKIN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PKIN activity, wherein PKIN is combined with at least one test compound, and the activity of PKIN in the presence of a test compound is compared with the activity of PKIN in the absence of the test compound. A change in the activity of PKIN in the presence of the test compound is indicative of a compound that modulates the activity of PKIN. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PKIN under conditions suitable for PKIN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PKIN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PKIN or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic*

Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential 5 therapeutic or toxic agents.

Polynucleotides encoding PKIN may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. 10 (1998) Science 282:1145-1147).

Polynucleotides encoding PKIN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PKIN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the 15 blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PKIN, e.g., by secreting PKIN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

### THERAPEUTICS

20 Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PKIN and human kinases. In addition, the expression of PKIN is closely associated with cancer, diseased, proliferative, cardiac, tumorous, and digestive tissues, degenerative diseases of the brain, suggesting that PKIN plays a role in necrotic disorders affecting the central nervous system, and neuronal tissues (e.g. brain and spinal cord, see Table 6). Therefore, PKIN 25 appears to play a role in maintenance and potentially the neoplastic transformation of cells of the central nervous system, and in cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders. In the treatment of disorders associated with increased PKIN expression or activity, it is desirable to decrease the expression or activity of PKIN. In the treatment of disorders associated with decreased PKIN expression or activity, it is 30 desirable to increase the expression or activity of PKIN.

Therefore, in one embodiment, PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins,

5       immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins,

10      erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus,

15      systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria,

20      polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia,

25      Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea

30      and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure,

35      ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PKIN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PKIN may be

administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those listed above.

In a further embodiment, an antagonist of PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN. Examples of such

5 disorders include, but are not limited to, those cancers, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders described above. In one aspect, an antibody which specifically binds PKIN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PKIN.

10 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate

15 therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

20 An antagonist of PKIN may be produced using methods which are generally known in the art. In particular, purified PKIN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PKIN. Antibodies to PKIN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and

25 fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PKIN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to

30 increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to

35 PKIN have an amino acid sequence consisting of at least about 5 amino acids, and generally will

consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PKIN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

5        Monoclonal antibodies to PKIN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and

10      Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PKIN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

20      Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for PKIN may also be generated. 25      For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

30      Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PKIN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PKIN epitopes is generally used, but a competitive binding assay may

also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PKIN. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of PKIN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PKIN epitopes, represents the average affinity, or avidity, of the antibodies for PKIN. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PKIN epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PKIN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PKIN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PKIN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PKIN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PKIN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PKIN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Cli. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995)

9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other 5 systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PKIN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency 10 (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene 15 Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites 20 (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the 25 case where a genetic deficiency in PKIN expression or regulation causes disease, the expression of PKIN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PKIN are treated by constructing mammalian expression vectors encoding PKIN and introducing these vectors by mechanical means into PKIN-deficient cells. Mechanical transfer technologies for 30 use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

35 Expression vectors that may be effective for the expression of PKIN include, but are not

limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PKIN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 5 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the 10 FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PKIN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver 15 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

20 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PKIN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PKIN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences 25 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. 30 (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by 35 reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4 $^{+}$  T-

cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver 5 polynucleotides encoding PKIN to cells which have one or more genetic abnormalities with respect to the expression of PKIN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be 10 versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242, both 15 incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver 20 polynucleotides encoding PKIN to target cells which have one or more genetic abnormalities with respect to the expression of PKIN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PKIN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. 25 Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. 30 (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

35 In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to

deliver polynucleotides encoding PKIN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PKIN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PKIN-coding RNAs and the synthesis of high levels of PKIN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of PKIN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PKIN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,

corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

5 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PKIN. Such DNA sequences may be incorporated into a wide variety of vectors  
10 with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3'  
15 ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterate linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous  
20 endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PKIN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides,  
25 transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PKIN expression or activity, a compound which specifically inhibits expression of the polynucleotide  
30 encoding PKIN may be therapeutically useful, and in the treatment of disorders associated with decreased PKIN expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PKIN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method  
35 commonly known in the art, including chemical modification of a compound known to be effective in

altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a

5 polynucleotide encoding PKIN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PKIN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence

10 of the polynucleotide encoding PKIN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide

15 can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide

20 nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells

25 taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of

30 such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins.

Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PKIN, antibodies to PKIN, and mimetics, agonists, antagonists, or inhibitors of PKIN.

The compositions utilized in this invention may be administered by any number of routes 5 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. 10 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides 15 and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of 20 macromolecules comprising PKIN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PKIN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to 25 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell 30 culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PKIN or fragments thereof, antibodies of PKIN, and agonists, antagonists or inhibitors of PKIN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by 35 calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose

lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is 5 preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the 10 active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

15 Normal dosage amounts may vary from about  $0.1\ \mu\text{g}$  to  $100,000\ \mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their 20 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PKIN may be used for the diagnosis of disorders characterized by expression of PKIN, or in assays to monitor patients being treated with PKIN or agonists, antagonists, or inhibitors of PKIN. Antibodies useful for diagnostic 25 purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PKIN include methods which utilize the antibody and a label to detect PKIN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

30 A variety of protocols for measuring PKIN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PKIN expression. Normal or standard values for PKIN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PKIN under conditions suitable for complex formation. The amount of standard complex formation may be 35 quantitated by various methods, such as photometric means. Quantities of PKIN expressed in subject,

control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PKIN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, 5 complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PKIN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PKIN, and to monitor regulation of PKIN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide 10 sequences, including genomic sequences, encoding PKIN or closely related molecules may be used to identify nucleic acid sequences which encode PKIN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PKIN, allelic variants, or related 15 sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PKIN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:19-36 or from genomic sequences including promoters, enhancers, and introns of the PKIN gene.

20 Means for producing specific hybridization probes for DNAs encoding PKIN include the cloning of polynucleotide sequences encoding PKIN or PKIN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a 25 variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PKIN may be used for the diagnosis of disorders 30 associated with expression of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress 35 syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis,

autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,

5 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner

10 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma,

15 leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy,

20 gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and

25 sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

30 degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung

35 anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary

hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases,

5   pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax,

10   pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and

15   ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous

20   xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding PKIN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PKIN expression. Such qualitative or

25   quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PKIN may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PKIN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a

30   suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PKIN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to

35   monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PKIN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PKIN, under conditions suitable for hybridization or amplification.

5 Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

10 Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

15 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development 20 or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PKIN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding PKIN, or a fragment of a polynucleotide complementary to the polynucleotide encoding 25 PKIN, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are 30 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, 35 from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause

differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis 5 methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry 10 using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PKIN include radiolabeling or biotinylation nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be 15 accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray 20 can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the 25 activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

30 In another embodiment, PKIN, fragments of PKIN, or antibodies specific for PKIN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of 35 gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of 5 transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, 10 biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental 15 compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share 20 those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for 25 comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at 30 <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of 35 the present invention, so that transcript levels corresponding to the polynucleotides of the present

invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present

5 invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by

10 separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by

15 staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the

20 spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

25 A proteomic profile may also be generated using antibodies specific for PKIN to quantify the levels of PKIN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a

30 variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor

35 correlation between transcript and protein abundances for some proteins in some tissues (Anderson,

N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such 5 cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological 10 sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological 15 sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated 20 sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-25 2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PKIN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either 30 coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), 35 yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1

constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the

5 inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man

10 (OMIM) World Wide Web site. Correlation between the location of the gene encoding PKIN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

*In situ* hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

15 Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23,

20 any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PKIN, its catalytic or immunogenic fragments, or

25 oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PKIN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds

30 having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PKIN, or fragments thereof, and washed. Bound PKIN is then detected by methods well known in the art. Purified PKIN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,

35 non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PKIN specifically compete with a test compound for binding PKIN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PKIN.

5 In additional embodiments, the nucleotide sequences which encode PKIN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding 10 description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

Without further elaboration, it is believed that one skilled in the art can, using the preceding 15 description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/199,021, U.S. Ser. No. 60/200,226, U.S. Ser. No. 60/202,339, U.S. Ser. No. 60/203,505, U.S. Ser. No. 60/205,654, U.S. Ser. No. 60/207,739, and U.S. Ser. No. 60/208,795, 20 are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database 25 (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium 30 acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was 35 isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA

purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the

MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides 5 were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques 10 disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public 15 databases such as the GenBank primate; rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The 20 Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on 25 GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, 30 and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program 35 (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide 10 and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:19-36. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative human kinases were initially identified by running the Genscan gene identification 15 program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a 20 FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode human kinases, the encoded polypeptides were analyzed by querying against PFAM models for kinases. Potential human kinases were also identified by homology to Incyte cDNA sequences that had been annotated as kinases. These selected Genscan-predicted sequences 25 were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, 30 this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

35 V. Assembly of Genomic Sequence Data with cDNA Sequence Data

**"Stitched" Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

**"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

**VI. Chromosomal Mapping of PKIN Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:19-36 were compared with

sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:19-36 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available 5 from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map 10 position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation 15 hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:24 was mapped to chromosome 2 within the interval from 92.30 20 to 103.1 centiMorgans, SEQ ID NO:25 was mapped to chromosome 11 within the interval from 104.8 to 117.9 centiMorgans, SEQ ID NO:33 was mapped to chromosome 8 within the interval from 25.8 to 40.3 centiMorgans, SEQ ID NO:23 was mapped to chromosome 9 within the interval from 101.20 to 104.90 centiMorgans, to chromosome 10 within the interval from 145.20 to 156.60 centiMorgans, and to chromosome 19 within the interval from 69.90 to 81.20 centiMorgans. More than one map 25 location is reported for SEQ ID NO:23, indicating that sequences having different map locations were assembled into a single cluster. This situation occurs when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a 30 gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is

much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

5

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \min\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is 10 calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate 15 the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% 20 identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PKIN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is 25 classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number 30 of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PKIN. cDNA sequences and cDNA library/tissue 35 information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### VIII. Extension of PKIN Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was 5 synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

10 Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

15 High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg<sup>2+</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94 °C, 3 min; Step 2: 20 94 °C, 15 sec; Step 3: 57 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C.

25 The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

30 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones 35 were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

5 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA 10 recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

15 In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:19-36 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base 20 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a 25 SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

30 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

35 **X. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), 5 *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., 10 Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The 15 array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of 20 complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and 25 poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with 30 GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

(CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

5 **Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 10  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR 15 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

20 Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

25 Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

**Hybridization**

30 Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash

buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

**Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an 5 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a 10 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, 15 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is 15 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on 20 the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the 25 two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high 30 signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then

integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### XI. Complementary Polynucleotides

Sequences complementary to the PKIN-encoding sequences, or any parts thereof, are used to 5 detect, decrease, or inhibit expression of naturally occurring PKIN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PKIN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent 10 promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PKIN-encoding transcript.

#### XII. Expression of PKIN

Expression and purification of PKIN is achieved using bacterial or virus-based expression systems. For expression of PKIN in bacteria, cDNA is subcloned into an appropriate vector 15 containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PKIN upon induction with isopropyl beta-D- 20 thiogalactopyranoside (IPTG). Expression of PKIN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PKIN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong 25 polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

30 In most expression systems, PKIN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham 35 Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from

PKIN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, 5 supra, ch. 10 and 16). Purified PKIN obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII where applicable.

### XIII. Functional Assays

PKIN function is assessed by expressing the sequences encoding PKIN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression 10 vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a 15 marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate 20 the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; 25 alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PKIN on gene expression can be assessed using highly purified populations 30 of cells transfected with sequences encoding PKIN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. 35 Expression of mRNA encoding PKIN and other genes of interest can be analyzed by northern analysis

or microarray techniques.

#### XIV. Production of PKIN Specific Antibodies

PKIN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to 5 immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PKIN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well 10 described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the 15 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for anti-peptide and anti-PKIN activity by, for example, binding the peptide or PKIN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XV. Purification of Naturally Occurring PKIN Using Specific Antibodies

20 Naturally occurring or recombinant PKIN is substantially purified by immunoaffinity chromatography using antibodies specific for PKIN. An immunoaffinity column is constructed by covalently coupling anti-PKIN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

25 Media containing PKIN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PKIN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PKIN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PKIN is collected.

#### 30 XVI. Identification of Molecules Which Interact with PKIN

PKIN, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PKIN, washed, and any wells with labeled PKIN complex are assayed. Data obtained using different concentrations 35 of PKIN are used to calculate values for the number, affinity, and association of PKIN with the

candidate molecules.

Alternatively, molecules interacting with PKIN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

5 PKIN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

## XVII. Demonstration of PKIN Activity

10 Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by PKIN in the presence of gamma-labeled  $^{32}\text{P}$ -ATP. PKIN is incubated with the protein substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the substrate is separated from free  $^{32}\text{P}$ -ATP by electrophoresis and the incorporated  $^{32}\text{P}$  is counted using a radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of PKIN. A 15 determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate 20 and gamma  $^{32}\text{P}$ -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated  $^{32}\text{P}$ -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma  $^{32}\text{P}$ -ATP. The reservoir of the centrifuged unit containing the  $^{32}\text{P}$ -peptide product as retentate 25 is then counted in a scintillation counter. This procedure allows assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes are as follows: Histone H1 (Sigma) and p34<sup>cdc2</sup>kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and 30 ERK (Pearson, J.D. et al. (1991) *Methods Enzymol.* 200:62-81).

In another alternative, protein kinase activity of PKIN is demonstrated *in vitro* in an assay containing PKIN, 50 $\mu$ l of kinase buffer, 1 $\mu$ g substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10  $\mu$ g ATP, and 0.5 $\mu$ Ci [ $\gamma$ - $^{32}\text{P}$ ]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [ $\gamma$ -

<sup>32</sup>P]ATP is removed by washing and the incorporated radioactivity is measured using a radioactivity scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and visualized on a 12% SDS polyacrylamide gel by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the 5 presence of the inactive kinase, K38A.

In yet another alternative, adenylate kinase or guanylate kinase activity may be measured by the incorporation of <sup>32</sup>P from gamma-labeled <sup>32</sup>P -ATP into ADP or GDP using a gamma radioisotope counter. The enzyme, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and <sup>32</sup>P-labeled ATP as the phosphate donor. The reaction 10 is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is cut out and counted. The radioactivity recovered is proportional to the enzyme activity.

In yet another alternative, other assays for PKIN include scintillation proximity assays (SPA), 15 scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of PKIN activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

### XVIII. Enhancement/Inhibition of Protein Kinase Activity

20 Agonists or antagonists of PKIN activation or inhibition may be tested using assays described in section XVII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in PKIN activity.

Various modifications and variations of the described methods and systems of the invention 25 will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the 30 scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Polynucleotide ID
2890544	1	2890544CD1	19	2890544CB1	
7472693	2	7472693CD1	20	7472693CB1	
3107952	3	3107952CD1	21	3107952CB1	
5544420	4	5544420CD1	22	5544420CB1	
7472832	5	7472832CD1	23	7472832CB1	
1551456	6	1551456CD1	24	1551456CB1	
2589355	7	2589355CD1	25	2589355CB1	
4357117	8	4357117CD1	26	4357117CB1	
5511992	9	5511992CD1	27	5511992CB1	
7474560	10	7474560CD1	28	7474560CB1	
7474602	11	7474602CD1	29	7474602CB1	
7475509	12	7475509CD1	30	7475509CB1	
7475491	13	7475491CD1	31	7475491CB1	
2192119	14	2192119CD1	32	2192119CB1	
7474496	15	7474496CD1	33	7474496CB1	
1834248	16	1834248CD1	34	1834248CB1	
71584520	17	71584520CD1	35	71584520CB1	
7475538	18	7475538CD1	36	7475538CB1	

Table 2

Polypeptide SEQ ID NO:	IncYTE Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	2890544CD1	g5305331	0	protein kinase Myak-L [Mus musculus]
2	7472693CD1	g790790	2.4e-144	cam kinase I [Homo sapiens]
3	3107952CD1	g1403532	2.3e-226	KIS protein kinase [Rattus norvegicus] (Maucuer,A. et al. (1997) J. Biol. Chem. 272: 23151-23156)
4	5544420CD1	g205278	6.4e-293	male germ cell-associated kinase (mak) [Rattus norvegicus] (Matsuhashi,H. et al. (1990) Mol. Cell. Biol. 10: 2261-2268)
5	7472832CD1	g438373	0.0	protein kinase C mu [Homo sapiens] (Johannes,F.J. et al. (1994) J. Biol. Chem. 269: 6140-6148)
6	1551456CD1	g4099088	3.8e-26	[Arabidopsis thaliana] SNF1 family protein kinase
7	2589355CD1	g6760436	1.2e-144	[Gallus gallus] gin-induced kinase
8	4357117CD1	g6552404	2.4e-199	[Rattus norvegicus] DLG6 alpha
9	5511992CD1	g971420	3.9e-231	mixed lineage kinase 2 [Homo sapiens] Dorow,D.S. et al. (1995) Eur. J. Biochem. 234:492-500
9	5511992CD1	g12005724	0	[5' incom] [Homo sapiens] mixed lineage kinase MLKL
10	7474560CD1	g4691541	7.3e-102	Adenylylate kinase 5 [Homo sapiens]. Van Rompay,A.R. et al. (1999) Identification of a novel human adenylylate kinase cDNA cloning, expression analysis, chromosome localization and characterization of the recombinant protein, Eur. J. Biochem. 261:509-516.
11	7474602CD1	g439614	8.9e-145	CaM-like protein kinase [Rattus norvegicus]. Cho, F.S. et al. (1994) Characterization of a rat cDNA clone encoding calcium/calmodulin-dependent protein kinase I, Biochim. Biophys. Acta 1224:156-160.

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
12	7475509CD1	g28577	2.4e-112	Nucleoside-triphosphate--adenylate kinase [Homo sapiens]. Xu, G. et al. (1992) Characterization of human adenylylate kinase 3 (AK3) cDNA and mapping of the AK3 pseudogene to an intron of the NF1 gene. <i>Genomics</i> 13:537-542.
13	7475491CD1	g2257588	1.4e-210	PCTAIRE3 [Rattus rattus]. Hirose, T. et al. (1997) PCTAIRE 2, a Cdc2-related serine/threonine kinase, is predominantly expressed in terminally differentiated neurons. <i>Eur. J. Biochem.</i> 249:481-488.
14	2192119CD1	g38880563	2.0e-121	Predicted using GeneFinder; similar to serine/threonine kinase; cDNA [Caenorhabditis elegans]. The C. elegans Sequencing Consortium (1998). <i>Science</i> 282:2012-2018.
14	2192119CD1	g10442581	0	105-kDa kinase-like protein [Mus musculus]. Liu, S.C.H., et al., (2000) <i>Biochim. Biophys. Acta</i> 1517:148-152.
15	7474496CD1	g60666585	0.0	GCN2 eIF2alpha Kinase [Mus musculus].
16	1834248CD1	g75955802	1.40E-252	ELKL motif kinase 2 short form [Mus musculus].
17	71584520CD1	g3927912	3.9e-157	calmodulin binding protein kinase [Fugu rubripes]. Cottage, A. et al. (1999) <i>FEBS Lett.</i> 443: 370-374.
18	7475538CD1	g4090958	9.2e-85	cell cycle related kinase [Homo sapiens].
18	7475538CD1	g9664926	0	CDK-related protein kinase PNQLARE [Mus musculus].

Table 3

SEQ ID NO:	Incyte ID Polypeptide ID:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2890544CD1	1210	S20 T107 T163 S211 T422 T666 S843 S853 T907 S127 T212 T508 S29 S37 T87 S113 S169 S211 S396 T441 T474 T643 S856 S910 T912 T938 S967 Y459 T1057 S1008 S1138 S1187	N57 N111 N133 N149 N262 N471 N566 N570 N1009 N1045	Protein Kinase Atp: L196-K219 Protein Kinase St: L311-L323 Eukaryotic protein kinase domain: Y190-P411 I492-V518 Tyrrosine kinase catalytic domain signature PD00109B: K305-L323 PROTEIN KINASE NUCLEAR HOMEO DOMAIN INTERACTING HOMEBOX DNABINDING SERINE/THREONINE PROTEIN: PD150874: G1030-L1210 PROTEIN KINASE DOMAIN: DM000004 [P14680] 371-694: V192-P509	MOTIFS MOTIFS HMMER_PFM BLIMPS_PRINTS BLAST_PRODOM
2	7472693CD1	357	S64 S102 T117 S154 S169 S251 S326 S343 Y116 Y133 S11 T45 T269	N225 N311 N332	PROTEIN KINASE DOMAIN DM00004 [S57347] 21-266: E24-C270 Protein Kinase Atp: L29-K52 Protein Kinase St: I140-Y152 signal cleavage: M1-A40 SPSCAN Eukaryotic protein kinase domain okinase: F23-I279 Protein kinases signatures and profile: D120-D176 Tyrrosine kinase catalytic domain signature PR00109A: M98-V111 PR00109B: Y134-Y152 PR00109D: V202-E224 PROTEIN KINASE CALMODULINBINDING I CALCIUM/ CALMODULIN DEPENDENT TYPE CAM PHOSPHORYLATION SERINE/THREONINE PROTEIN EDD012137 : W278-L322	BLIMPS_PRINTS BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	3107952CD1	419	S67 S117 S181 S215 S221 T244 S290 T390	N253	transmembrane domain: V223-L241 Eukaryotic protein kinase domain: W23-F304	HMMER_HMMER_PFFAM BLIMPS_PRINTS

Table 3 (cont.)

SEQ NO:	Incyte ID Polyptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	5544420CD1	624	T274 S424 S436 S496 S41 S179 S204 S315 S406 S451 S466 S500 S530 T6 S53 S161 T218 S388 S391 S420 S463 T475 T581 S587 T594 Y15 Y76 T571	N24 N252 N352 N384 N449 N455 N543 N569	Biakaryotic protein kinase domain Y4-F284 Tyrosine kinase catalytic domain PR00109B:F115-C133 PR00109D:I181-T203 PR00109E:A253-A275 Protein Kinases signatures & profile: V101-Q153 Protein Kinases ATP-binding region: L10-K33	BLIMPS_PRINTS HMMER_PFAM
5	7472832CD1	878		S197 S396 S797 S75 S111 T130 T218 S225 S333 S353 S362 T398 S408 T573 T589 T812 T853 S145 S206 T247 S387 T391 T392 T412 T434 S604 S641 S706 Y87	N187 N431 N432 N454 N473 N729 H139-C188 H265-C314 Eukaryotic protein kinase domain: I151-L807 Phorbol esters/diacylglycerol binding PH domain: T398-V478 Protein Kinases signatures & profile: L650-S706	BLIMPS_PRINTS HMMER_PFAM HMMER_PFAM HMMER_PFAM PROFILESCAN

Table 3 (cont.)

SEQ NO:	Incyte ID Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5					Phorbol esters/diacylglycerol binding F151-S214 C278-A339	PROFILESCAN
					Tyrosine kinase catalytic domain: PR00109B:H664-L682 PR00109D:L732-D754	BLIMPS_PRINTS
					Diacylglycerol/phorbol-esterase PR0008B:C152-G161 PR0008C:Q291-C302 PR0008D:H303-L315	BLIMPS_PRINTS
					KINASE PHORBOLESTER C MU SERINE/THREONINE NPKCMU ATP BINDING PD039353:D323-V478 PD031784:V45-P138 PD027900:L807-L878	BLAST_PRODOM
					KINASE PHORBOLESTER BINDING TRANSFERASE SERINE/THREONINE ZINC ATPBINDING C DUPLICATION PD000215:H265-C314	BLAST_PRODOM
					Protein kinases signatures and profile (protein kinase_tyr_prf): R826-A877	BLAST_PRODOM
					PROTEIN KINASE DOMAIN DM00004 A53215 585-829:B553-V798 DM00004 I48719 591-835:P553-V798 DM04692 A37237 1-676:V556-S771 M04692 P05773 1-672:V556-P845	BLAST_DOMAIN
					Phorbol esters / diacylglycerol binding domain: H139-C188 H265-C314	MOTIFS
					Protein kinases ATP-binding region signature: L557-K580	MOTIFS
					Serine/Threonine protein kinases active-site signature: I670-L682	MOTIFS

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	1551456CD1	440	S11 S175 T74 S85 S230 T297 S361 S365 S376 S396 T426 S11 S19 S131 S155 T214 S251 Y68 Y378		Eukaryotic protein kinase domain: L1164-E331 Tyrosine kinase signature: AL92-L210, S260-S282 PROTEIN KINASE DOMAIN DM00004/P34244   82-359: Y167-L324	HHMER -PFAM BLIMPS -PRINTS BLAST -DOMO
7	2589355CD1	923	Y104 Y676 T258 S355 T481 S584 T38 S125 S289 T356 S391 T447 S448 T455 T481 S500 S512 T875 S890 T17 S254 S336 T417 S472 S531 S542 S551 S564 S576 S667	N112 N317	PROTEIN KINASE DOMAIN DM00004/P27448   58-297: G25-I259 Serine/Threonine protein kinases active-site signature Protein Kinase St: I135-L147 Eukaryotic protein kinase domain Kinase: S24-M268 Protein kinases signatures and profile Protein kinase_tyr.prf: Y90-G167 Tyrosine kinase catalytic site PR00109: T93-A106, Y129-L147, L195-P217	BLAST -DOMO MOTIFS MOTIFS HHMER -PFAM BLIMPS -PRINTS
8	4357117CD1	442	T16 T267 T270 S295 S361 S142 S229 S178 S180 T311 S408 S437 Y294 Y304 Y346	N406	SH3 domain signature PR00452: C147-R159, A115-Q130, D132-I141 PROTEIN DOMAIN SH3 KINASE GUANYLATE TRANSFERASE ATPBINDING REPEAT GMP MEMBRANE PD001338: T267-Q360 GUANYLATE KINASE DM00755/A57653   370- 570: P228-P431 PDZ signaling molecule domain PDZ: I3-V83 Guanylate kinase Guanylate_kin: T268-Y372 Guanylate kinase protein BL00856: V264-Y284, L292-R339	BLIMPS -PRINTS BLAST -PRODOM BLAST -PRODOM BLAST -DOMO HHMER -PFAM HHMER -PFAM BLIMPS -BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	5511992CD1	1046	T72 T112 S118 S223 S276 T294 S603 S707 S770 T781 S814 S880 T913 S531 T777 S89 T135 T363 T394 T395 T436 S599 S606 S636 T644 S808 S821 S834 T884 T919 S923 S966 S972 S983 Y984 Y325	N813 N862 N964	Receptor tyrosine kinase BL00240: E290-V337, V337-I389 Receptor tyrosine kinase BL00239: E181-P228, L232-I254, W291-R340, L345-I389 Receptor tyrosine kinase BL00790: I154-C207, S298-W330, L356-M404 Protein kinase signature and profile PROFILESCAN Protein kinase_tyr.prf: L232-T294 Tyrrosine kinase catalytic site PR00109: N210-S223, D248-I266, G301-I311, S320-I342, C364-F386 SH3 domain signature PR00452: P55-A65, D69-K84, D91-N100, R102-R114 SH3 domain: P55-R114 Eukaryotic protein kinase domain HMMER-PFAM Kinase: L134-L393 Kinase domain SH3 MIXED LINEAGE BLAST-PRODOM	BLIMPS-BLOCKS
					SERINE/THREONINE WITH LEUCINE ZIPPER PROLINE PD024997: I396-A741 PROTEIN KINASE DOMAIN DM00004/A53800 119-368: L136-F386 Protein kinases ATP-binding region signature Protein_Kinase_Atp: I140-K161 Serine/Treonine protein kinases active-site signature MOTIFS Protein Kinase St: I254-I266	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incycyte ID Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7474560CD1	357	S23, T65, Y103, S124, T140, S142, S181, T185, S200, S211, S213, S245, T288, T304, Y316, T340, T350		KINASE ADENYLYLATE TRANSFERASE, ATP-BINDING ATP/AMP TRANSPHOSPHORYLASE ISOENZYME PROTEIN 3D-STRUCTURE MITOCHONDRION: PDD000657: I175-R294. ADENYLYLATE KINASE: DM00290   P00570   1-131: E165-R294.	BLAST-PRODOM
11	7474602CD1	355	S11, T45, S64, S102, Y116, T117, Y133, S154, S169, S251, T269, S326, S343	N225, N311, N332	Adenylylate kinase protein: BL00113A: F174-L190; BL00113B: H198-A241; BL00113C: G247-V261; BL00113D: S298-T328. Adenylylate kinase signature: PR00094A: F174-C187; PR00094B: G202-S216; PR00094C: F252-G268; PR00094D: P300-Y315; PR00094E: R317-Q331. Uridine kinase signature: PR00988A: C170-C187	BLIMPS-BLOCKS BLIMPS-PRINTS
					Shikimate kinase family: PR01100A: I173-E188.	BLIMPS-PRINTS
					Adenylylate kinase: I9-Q20, F252-Q263.	MOTIFS
					Adenylylate kinase: I11-L88, I175-Q331.	HMMER-PFAM
					Adenylylate kinase signature (adenylylate kinase.prf): V229-A284.	PROFILESCAN
					PROTEIN KINASE DOMAIN: DM0004   S57347   21-266: E24-C270.	BLAST-PRODOM
					Protein kinases ATP-binding region signature: L29-K52.	MOTIFS
					Serine/Threonine protein kinases active-site signature: I140-Y152.	MOTIFS
					Eukaryotic protein kinase domain: F23-I279.	HMMER-PFAM
					Protein kinases signatures and profile (protein kinase_tyr.prf): D120-D176.	PROFILESCAN
					Tyrosine kinase catalytic domain: PR00109A: M98-V111; PR00109B: Y134-Y152; PR00109D: V202-E224.	BLIMPS-PRINTS
					PROTEIN KINASE; CALMODULINBINDING I CALCIUM/CALMODULINDEPENDENT TYPE CAM TRANSFERASE SERINE/THREONINE PROTEIN PHOSPHORYLATION: PD012137: W278-L322.	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte ID Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7475509CD1	224	S16, S75, T118, S124, S132, T201,		ADENYLYLATE KINASE: DM00290   P27144   1-125: M1-R126.	BLIMPS-PRINTS
					Adenylylate kinase signature: PR00094A: V9-C22; PR00094B: G37-G51; PR00094C: W85-D101; PR00094D: Q159-Y174; PR00094E: D176-V190.	BLIMPS-PRINTS
13	7475491CD1	502	S12, T20, S24, S60, S64, S87, T104, S122, S126, S156, S160, Y183, T232, T239, T348, T364, T373, T387, S455, S471, S500,		Shikimate kinase family: PRO1100A: A8-Q23; PRO1100E: D107-S124 (p<0.0028). Adenylylate kinase Motif: W85-Q96.	BLIMPS-PRINTS
					KINASE ADENYLYLATE TRANSFERASE, ATP-BINDING ATP/AMP TRANSPHOSPHORYLASE ISOENZYME, PROTEIN 3D-STRUCTURE MITOCHONDRION: PD000657: I10-V190.	MOTIFS
					Adenylylate kinase: I10-V190	BLAST-PRODOM
					Adenylylate kinase protein: BL00113A: V9-I25; BL00113B: H33-E76; BL00113C: R80-L94; BL00113D: S132-D162.	BLIMPS-BLOCKS
					Adenylylate kinase signature (adenylylate kinase.prf): V64-F116.	HMMER-PFAM
					PROTEIN KINASE DOMAIN: DM0004   Q04899   122-392: V173-A444.	BLAST-DOMO
					KINASE SERINE/THREONINE PROTEIN TRANSFERASE ATP-BINDING DOMAIN - PCTAIRE1, PCTAIRE2, PCTAIRE3, CRK5 ALTERNATIVE: PD007333: D120-T171	BLAST-PRODOM
					Tyrosine kinase catalytic domain: PRO109B: Y283-T301	BLIMPS-PRINTS
					Protein kinases signatures and profile (protein_kinase_tyr.prf): D235-A317	PROFILESCAN
					Eukaryotic protein kinase domain (kinase): Y172-F453.	HMMER-PFAM
					Protein kinases ATP-binding region signature: L178-K201	MOTIFS
					Serine/Threonine protein kinases active-site signature: I289-T301	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte ID	Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Protein Kinase Domain	Analytical Methods and Databases
14	2192119CD1	791		T39, T85, S193, S269, T319, S342, S359, T361, T408, S462, S499, S521, S535, S539, T545, T581, T610, T617, S622, S662, S668, S677, S693, S716, S726, S737, Y184	N447	PROTEIN KINASE DOMAIN DM00004 P52304  27-267: K69-P254 (p > 1.3e-06)		BLAST-DOMO

Table 3 (cont.)

SEQ ID NO: 15	Incyte Polypeptide ID: 7474496CD1	Amino Acid Residues 1651	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			T65, S108, S210, S216, S250, T346, T409, S414, S450, S553, T581, S690, S709, T715, S722, S733, S755, S757, T811, T824, S882, S917, T945, S947, S961, T993, T1022, S1038, S1059, T1062, T1182, T1234, T1238, S1397, T1418, S1437, S1457, S1572, T1475, S1547, T1548, T1604, S1641, Y256, Y821, Y840, Y1626	N100, 1057, 1203, 1248, 1416, 1520, 1602	PROTEIN KINASE DOMAIN: DM00004 P15442 645-902: S732-A994, EUKARYOTIC INITIATION FACTOR KINASE (ELF2 ALPHA): P0156018: D13-D219 Tyrosine kinase catalytic domain: ER00109B: Y840-L858 Protein kinases signatures and profile (protein kinase_tyr.prf): R826-A877	BLAST-DOMO BLAST-PRODOM BLIMPS-PRINTS
					Eukaryotic protein kinase domain (protein kinase_tyr.prf): R826-A877	PROFILESCAN
					K335-D446; F592-P664; Y799-L1003	HMMER-PFAM
					Protein kinases ATP-binding region	MOTIFS
					L598-K621	MOTIFS
					Serine/Threonine protein kinases active-site signature: M846-L858	MOTIFS
					Eukaryotic protein kinase domain: K335-D446; P504-I541; F592-P664; Y799-L1003	HMMER-PFAM
					Protein kinases signatures and profile (protein kinase_tyr.prf): R826-A877	PROFILE-SCAN
					Tyrosine kinase catalytic domain: ER00109B: Y840-L858	BLIMPS-PRINTS
					Protein kinases ATP-binding region signature: L598-K621	MOTIFS
					Serine/Threonine protein kinases active-site signature: M846-L858	MOTIFS

Table 3 (cont.)

SEQ ID NO.	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Potential Glycosylation	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	1834248CD1	752	S139 S210 S23 S27 S34	N395 N532		Eukaryotic protein kinase domain: Y59-L130,	HMMER_PFFAM
			S354 S399 S423 S441 S458 S48			Tyrosine kinase catalytic domain	ELIMPS_PRINTS
			S494 S661 S666			PR00109A: M135-V148	
			S710 T127 T281			PR00109B: Y171-L189	
			T300 T323 T332			PR00109D: V237-H259	
			T344 T507 T511			Protein kinases signatures and profile protein_kinase_tyr.prf: Y132-S210	PROFILESCAN
			T515 T536 T564			Protein kinases ATP-binding region: Protein_Kinase_Atp: I65-K88	MOTIFS
			T620 T624 T81			Serine/Threonine protein kinases active-site Protein_Kinase_St: I177-L189	MOTIFS
						KINASE SERINE/THREONINE TRANSFERASE ATP BINDING PROTEIN EMK P78 CDC25C	BLAST_PRODOM
						PD008571: S412-E632	
						KINASE SERINE/THREONINE TRANSFERASE ATP BINDING PROTEIN PAR1 KP78 EMK	BLAST_PRODOM
						PD005838: I310-R410	
						KINASE SERINE/THREONINE TRANSFERASE ATP BINDING KIN1 EMK PAR1	BLAST_PRODOM
						PD004300: E650-L752	
						KINASE TRANSFERASE ATP BINDING SERINE/THREONINE PHOSPHORYLATION RECEPTOR TYROSINE	BLAST_PRODOM
						PD000001: Y59-Y137	
						PROTEIN KINASE DOMAIN	BLAST_DOMO
						DM00004 P27448 58-297:L61-L301	
						DM00004 I48609 55-294:L61-L301	
						DM00004 Q05512 55-294:L61-L301	
						DM00004 JC1446 20-261:R60-L301	

Table 3 (cont.)

SEQ NO:	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	71584520CD1	501	S118 S341 S364 S482 S483 S495 T103 T21 T276 T422 T46 T470 T51 T7 T91 Y135 Y491	S138 S292 S341 S364 S482 S483 S495 T103 T21 T276 T422 T46 T470 T51 T7 T91 Y135 Y491	Eukaryotic protein kinase domain kinase: E37-I286 Tyrosine kinase catalytic domain PR00109B: Y135-Y153 PR00109D: V201-E223 CALMODULIN BINDING PROTEIN PD059862 : G368-V443	BLIMPS_PRINTS HMMER_PFM

Table 3 (cont.)

SEQ ID NO:	Incyte ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7475538CD1	346	S241 Y176 Y215		Bukaryotic protein kinase domain pkinase: Y4-F288.	Tyrosine kinase catalytic domain PR00109B: F117-I135	BLIMPS_PRINTS HMMER_PFM
					Protein kinases signatures and profile protein kinase_eyr.prf: A69-D155	Protein kinases signatures and profile protein kinase_eyr.prf: A69-D155	PROFILESCAN
					Protein kinases ATP-binding region signature Protein_Kinase_Atp: I10-K33	MOTIFS	
					Serine/Threonine protein kinases active-site Protein_Kinase_St: I123-I135	MOTIFS	
					KINASE TRANSFERASE SERINE/THREONINE ATP BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN PD002608: V164-F288	BLAST_PRODOM	
					KINASE PROTEIN TRANSFERASE ATP BINDING SERINE/THREONINE PHOSPHORYLATION RECEPTOR TYROSINE TRANSMEMBRANE PD00001: Y169-P301	BLAST_PRODOM	
					PROTEIN KINASE DOMAIN DM00004 P29620 I21-289:I10-A279	BLAST_DOMO	
					DM00004 Q00526 6-286:R9-F288		
					DM00004 P43450 6-276:R9-A279		
					DM00004 P23437 6-286:R9-F288		

Table 4

Poly-nucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
19	2890544CB1	4224	1-1378, 4153-4224	6474032H1 (PLACFEB01)	545	1204
				71089659V1	2981	3638
				710833920V1	2936	3576
				71083254V1	2263	2853
				71084605V1	2351	2959
				7286979H1 (BRAIFFER06)	1	408
				71254276V1	1091	1695
				5980233F7 (MCOLDXT02)	3498	4224
				71252923V1	1632	2197
				7086636H1 (BRAUTDR03)	222	702
				71082857V1	1680	2356
				7313223H1 (LIVRREE02)	498	928
20	7472693CB1	1736	72-149, 1337-1736	70520498D1	760	1354
				609792R6 (COLNNOT01)	145	736
				91544947	1	385
				70518493D1	716	1182
				70518085D1	1172	1736
21	3107952CB1	1824	1802-1824	7308879H1 (MLR1DT01)	1215	1807
				2925973F6 (TLYMMNOT04)	715	1215
				3459433F6 (293TF1T01)	1	582
				2927552F6 (TLYMMNOT04)	1309	1824
				3107952F6 (BRSTTUT15)	507	1161
22	5544420CB1	2201	1285-1629, 2128-2201, 256-376	6922389H1 (PLACFER06)	1458	1597
				7739285H1 (THYMMOE01)	646	834
				5544420F6 (TESTNOC01)	807	1187
				4206166F6 (BRONNOT02)	2041	2201
				GNN:95924006_004.edit	368	1871
				GBI:95924006_raw.compo	1	492
				2512558F6 (LIVRUT04)	1617	1720
				92882961	1644	2127
				6909108J1 (BITTDIR01)	1138	1317
				6910588J1 (BITTDIR01)	1994	2577
				2940460H1 (THYMEET02)	2682	2956
				7274331H2 (KIDETXJ01)	1876	2503
23	74728332CB1	2974	2933-2974, 1-654	60205600U1	559	1091

Table 4 (cont.)

Poly-nucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
23				60205598U1 (LJDNGNON07)	690	1339
				6426807H1 (SPLNNNOT09)	2411	2951
				3344032H1 (SKIRNOR01)	2708	2958
				6811278U1 (SKIRNOR01)	1	647
				6910588H1 (PITUDIROL1)	1369	2003
				91860144	2512	2974
24	1551456CB1	3648	1906-2052, 1-752	1732303F6 (BRSTUTU08)	1230	1731
				6830947U1 (SINTNOR01)	2357	3121
				6834559H1 (BRSTNNON02)	1610	2330
				6808991U1 (SKIRNOR01)	222	921
				1389125H1 (EOSINNOT01)	1	246
				1680356F6 (STOMFET01)	3139	3648
				6870264H1 (BRAGNON02)	3374	3648
				7714085J1 (SINTTFFEE02)	298	950
				7746056H1 (ADRETUE04)	875	1491
				3296183E6 (TLXJINT01)	1553	2310
				70349297D1	3030	3648
				70872871V1	953	1599
				70875814V1	2218	2955
25	2589355CB1	4719	1835-1863, 1-1135, 2380-3186	6500833H1 (PROSTUS25)	2498	3226
				2313925T6 (NGANNOT01)	3264	3885
				6483636F9 (MIXDUNB01)	2716	3270
				70985467V1	388	980
				7251743H1 (PROSTMY01)	810	1299
				6864125H1 (BRAGNON02)	1323	1998
				6056925H1 (BRAENOT04)	1322	1735
				6705826H1 (HEAADIR01)	4080	4719
				7652463H2 (STOMTDE01)	997	1704
				7039646H1 (UTRSTMNR02)	1	566
				7257374H1 (SKIRTD01)	1977	2560
				70402344D1	3105	3674
				2586850F6 (BRAITUT22)	3840	4348
				924331H1 (RATRNOT02)	4009	4386
26	4357117CB1	1651	1-267	GS:4357117.Fasta	1	1329
				6131509F6 (BMARTXT02)	1134	1651

Table 4 (cont.)

Poly-nucleotide SEQ ID NO:	Imocyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
27	5511992CB1	3141	1672-1833, 71-157, 665-828, 1298-1538, 2580-2878	GS_5511992.fasta	1	3141
28	7474560CB1	1244	1205-1244, 1-131	6311370H1 (NERDTDN03)	616	1236
29	7474602CB1	1661	72-149	6853555H1 (BRAIFFEN08)	1	683
30	7475509CB1	912	881-912	6997205H1 (BRAXTD17)	691	1244
31	7475491CB1	2858	1-269, 2768-2858	70518523D1	577	1171
32	2192119CB1	2817	1-249	6772112J1 (BRAUNOR01)	1	640
33	7474496CB1	5305	3463-3753, 1403-1781, 2384-2464, 1-644, 5126-5305	346275T7 (THYMN0T02)	1043	1661
				6124350H1 (BRAHN0N05)	634	1180
				944796H1 (RATRNN0T02)	1	93
				3616204H6 (EPIPN0T01)	69	699
				71147872V1	193	912
				7232424H1 (BRAXTD15)	1831	2429
				7004062H1 (COLNFE01)	1967	2495
				7677070J1 (NOSETUE01)	1	564
				7716833J1 (SINTFEE02)	1293	1877
				754239R6 (BRAITUT02)	1000	1575
				7067423H1 (BRAUNOR01)	499	1106
				7428450H1 (UTRMTMR02)	2179	2858
				70680595V1	654	1111
				1687835F6 (PROSTUT01)	2416	2817
				6769461J1 (BRAUNOR01)	221	385
				1819105F6 (PROSN0T20)	1052	1721
				7716523H1 (SINTPEE02)	498	1072
				1796441T6 (PROSTUT05)	2186	2793
				1618475T6 (BRAITUT12)	2117	2783
				7675143H1 (NOSETUE01)	1130	1739
				6145288H1 (BRANDIT03)	1543	2170
				31896653H1 (THYMN0N04)	1855	2196
				7646563H1 (UTRSTUE01)	367	1049
				96700560	1	393
					3332	3875

Table 4 (cont.)

Poly-nucleotide SEQ ID NO:	Incycyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
33				429360R6 (BLADNOT01)	2735	3311
				70885734V1	4100	4744
				1830377F6 (THP1AZT01)	3649	4093
				6488464H1 (MIXDUNB01)	2852	3473
				488190R6 (HNT2AGT01)	2317	2865
				70888476V1	4651	5305
34	1834248CB1	3269	1754-1773, 1-142, 3072-3269	1832566R6 (BRAINNON01)	4034	4547
				GNN.G7139831_000025_004	1293	2275
				6146293H1 (BRANDIT03)	1791	2354
				6272238H2 (BRAIFEN03)	2635	3263
				6893004J1 (BRAITDRO3)	684	1248
				7663341J1 (UTRSTIME01)	873	1492
				4001427R6 (HNT2AZS07)	2927	3264
				60202068B1	2368	2916
				60202069B1	2334	2864
				6954283H1 (BRAITDRO2)	1	718
35	71584520CB1	3017	2968-3017, 1-49, 1234-1322, 679-736	9810284	2771	3269
				1287320T6 (BRAINOT11)	2615	2983
				71579751V1	1083	1791
				71580777V1	586	1149
				6764749H1 (BRAUNOR01)	1406	2001
				7581090H1 (BRAIFEC01)	1	577
				1295833H1 (PGANNOT03)	2783	3017
				1414795F6 (BRAINOT12)	2183	2508
				7362189H1 (BRAIFEE05)	458	1003
				215711216 (BRAINOT09)	2388	2977
				2157112F6 (BRAINOT09)	1963	2450
36	7475538CB1	2168	811-870, 933-1227	6855691H1 (BRAIFEN08)	1500	2168
				70644867V1	613	1261
				70645804V1	515	1249
				70645323V1	1	595
				71564044V1	1277	1993
				71565564V1	1156	1924

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
19	2890544CB1	MCLDNTXT02
20	7472693CB1	COLNNOT01
21	3107952CB1	TLYMNNOT04
22	5544420CB1	SEMVNOT01
23	7472832CB1	SKTRNOR01
24	1551456CB1	LVENNOT03
25	2589355CB1	BRADDI01
26	4357117CB1	BMARTXT02
27	5511992CB1	SINTFEE02
28	7474560CB1	BRAYDINO3
29	7474602CB1	COLNNOT01
30	7475509CB1	BRATUT21
31	7475491CB1	SCOMDIT01
32	2192119CB1	PROSTUS23
33	7474496CB1	BRAINNON01
34	1834248CB1	BRATUT22
35	71584520CB1	BRAIFEC01
36	7475538CB1	BRAIFEN08

Table 6

Library	Vector	Library Description
BMARTXT02	PINCY	Library was constructed using RNA isolated from treated SH-SY5Y cell line derived from bone marrow neuroblastoma tumor cells removed from a 4-year-old Caucasian female. The cells were cultured in the presence of retinoic acid.
BRADDIR01	PINCY	Library was constructed using RNA isolated from the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.
BRAFFEC01	PINCY	This large size-fractionated library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAFFEN08	PINCY	This normalized fetal brain tissue library was constructed from 400 thousand independent clones from a fetal brain tissue library. Starting RNA was made from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAINTON01	PSPORT1	Library was constructed and normalized from 4.88 million independent clones from a brain tissue library. RNA was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeval lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
BRATTUT21	PINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningothelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
BRATTUT22	PINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the right frontal-parietal lobe of a 76-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a meningioma. Family history included senile dementia.

Table 6 (cont.)

Library	Vector	Library Description
BRAYDIT03	PINCY	This normalized brain tissue library was constructed from 6.7 million independent clones from the BRAYDIT01 tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones.
COLINNOT01	PSPORT1	Library was constructed using RNA isolated from colon tissue removed from a 75-year-old Caucasian male during a hemicolectomy.
LYENNNOT03	PSPORT1	Library was constructed using RNA isolated from the left ventricle tissue of a 31-year-old male.
MCLDITXT02	PINCY	Library was constructed using RNA isolated from treated umbilical cord blood dendritic cells removed from a male. The cells were treated with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), stem cell factor (SCF), phorbol myristate acetate (PMA), and ionomycin. The GM-CSF was added at time 0 at 100 ng/ml, the TNF alpha was added at time 0 at 2.5 ng/ml, the SCF was added at time 0 at 25 ng/ml. The PMA and ionomycin were added at 13 days for five hours. Incubation time was 13 days.

Table 6 (cont.)

Library	Vector	Library Description
PROSTUS23	PINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.
SCOMDIT01	PINCY	Library was constructed using RNA isolated from diseased spinal cord tissue removed from the base of the medulla of a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema.
SEMVNOT01	PINCY	Library was constructed using RNA isolated from seminal vesicle tissue removed from a 58-year-old Caucasian male during radical prostatectomy. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 3+2) of the prostate. Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Family history included a malignant breast neoplasm.
SINTFEE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation. Serology was negative.
SKIRNCR01	PCDNA2.1	This random primed library was constructed using RNA isolated from skin tissue removed from the breast of a 17-year-old Caucasian female during bilateral reduction mammoplasty. Patient history included breast hypertrophy. Family history included benign hypertension.
TLYMNCT04	PINCY	Library was constructed using RNA isolated from activated Th1 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and B7-transfected COS cells, and then activated for six hours with anti-CD3 and anti-CD28 antibodies.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) <i>J. Mol. Biol.</i> 215:403-410; Altschul, S.F. et al. (1997) <i>Nucleic Acids Res.</i> 25:3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Assembled ESTs</i> : fasta Identity= 95% or greater and <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, fastx, tfastrx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) <i>Proc. Natl. Acad. Sci. USA</i> 85:2444-2448; Pearson, W.R. (1990) <i>Methods Enzymol.</i> 183:63-98; and Smith, T.F. and M.S. Waterman (1981) <i>Adv. Appl. Math.</i> 2:482-489.	<i>ESTs</i> : fasta E value= 1.0E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value= 1.0E-8 or less <i>Full Length sequences</i> : fasta score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) <i>Nucleic Acids Res.</i> 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) <i>Methods Enzymol.</i> 266:88-105; and Attwood, T.K. et al. (1997) <i>J. Chem. Inf. Comput. Sci.</i> 37:417-424.	Probability value= 1.0E-3 or less
HMmer	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) <i>J. Mol. Biol.</i> 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) <i>Nucleic Acids Res.</i> 26:320-322; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nutshell, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Barroch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCGr-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Person, B. and P. Argos (1994) <i>J. Mol. Biol.</i> 237:182-192; Person, B. and P. Argos (1996) <i>Protein Sci.</i> 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) <i>Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol.</i> , Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Barroch, A. et al. (1997) <i>Nucleic Acids Res.</i> 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18,
  - b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2-18,
  - c) a naturally occurring polypeptide comprising an amino acid sequence at least 98% identical to an amino acid sequence of SEQ ID NO:1,
  - d) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and
  - e) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.
- 15 2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-18.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
- 20 4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:19-36.
- 25 6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
- 30 8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide

comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

- b) recovering the polypeptide so expressed.

5 10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,

10 b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-36,

c) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 98% identical to the polynucleotide sequence of SEQ ID NO:19,

d) a polynucleotide complementary to a polynucleotide of a),

15 e) a polynucleotide complementary to a polynucleotide of b),

f) a polynucleotide complementary to a polynucleotide of c), and

g) an RNA equivalent of a)-f).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a  
20 polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

25 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and  
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

30

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

5

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected 10 from the group consisting of SEQ ID NO:1-18.

18. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition of claim 16.

15

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of 25 functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

30

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 5 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

10

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- 15 b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a 20 compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 25 a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

30

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific 35 hybridization complex is formed between said probe and a target polynucleotide in the biological

sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of PKIN in a biological sample comprising the steps of:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

15

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a  $F(ab')_2$  fragment, or
- e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

25 32. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

30

34. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibodies from said animal; and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

10

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

15 38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibody producing cells from the animal;
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;
- d) culturing the hybridoma cells; and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

25 39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

30

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant

immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in a sample, comprising the steps of:

- 5        a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in the sample.

10

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 from a sample, the method comprising:

- 15        a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

20        46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

25        49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

30        51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
- 10 58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
- 15 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
- 20 63. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:19.
64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.
- 25 65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:21.
66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:22.
- 30 67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:23.

68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:24.

69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 5 NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:26.

10 71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 15 NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.

74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 20 NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 25 NO:31.

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 30 NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID 35 NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

NO:36.

<110> INCYTE GENOMICS, INC.  
YUE, Henry  
GANDHI, Ameena R.  
TRIBOULEY, Catherine M.  
KEARNEY, Liam  
GRIFFIN, Jennifer A.  
NGUYEN, Dannie B.  
BANDMAN, Olga  
LU, Dyung Aina M.  
LAL, Preeti  
BURFORD, Neil  
KHAN, Farrah A.  
WALIA, Narinder K.  
YAO, Monique G.  
PATTERSON, Chandra  
BURRILL, John D.  
MARCUS, Gregory A.  
ZINGLER, Kurt A.  
RECIPON, Shirley A.  
LU, Yan  
POLICKY, Jennifer L.  
THORNTON, Michael  
TANG, Y. Tom  
HAFALIA, April  
ELLIOTT, Vicki S.  
BAUGHN, Mariah R.  
WALSH, Roderick T.  
RAMKUMAR, Jayalaxmi  
BOROWSKY, Mark L.  
AU-YOUNG, Janice  
HILLMAN, Jennifer L.  
GURURAJAN, Rajagopal

<120> HUMAN KINASES

<130> PI-0076 PCT

<140> To Be Assigned  
<141> Herewith

<150> 60/199,021; 60/200,226; 60/202,339; 60/203,505; 60/205,654;  
60/207,739; 60/208,795  
<151> 2000-04-20; 2000-04-28; 2000-05-05; 2000-05-11; 2000-05-18;  
2000-05-26; 2000-06-01

<160> 36

<170> PERL Program

<210> 1

<211> 1210

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2890544CD1

<400> 1

Met Ala Ser Gln Leu Gln Val Phe Ser Pro Pro Ser Val Ser Ser  
1 5 10 15  
Ser Ala Phe Cys Ser Ala Lys Lys Leu Lys Ile Glu Pro Ser Gly  
20 25 30

Trp Asp Val Ser Gly Gln Ser Ser Asn Asp Lys Tyr Tyr Thr His  
 35 40 45  
 Ser Lys Thr Leu Pro Ala Thr Gln Gly Gln Ala Asn Ser Ser His  
 50 55 60  
 Gln Val Ala Asn Phe Asn Ile Pro Ala Tyr Asp Gln Gly Leu Leu  
 65 70 75  
 Leu Pro Ala Pro Ala Val Glu His Ile Val Val Thr Ala Ala Asp  
 80 85 90  
 Ser Ser Gly Ser Ala Ala Thr Ser Thr Phe Gln Ser Ser Gln Thr  
 95 100 105  
 Leu Thr His Arg Ser Asn Val Ser Leu Leu Glu Pro Tyr Gln Lys  
 110 115 120  
 Cys Gly Leu Lys Arg Lys Ser Glu Glu Val Asp Ser Asn Gly Ser  
 125 130 135  
 Val Gln Ile Ile Glu Glu His Pro Pro Leu Met Leu Gln Asn Arg  
 140 145 150  
 Thr Val Val Gly Ala Ala Ala Thr Thr Thr Thr Val Thr Thr Lys  
 155 160 165  
 Ser Ser Ser Ser Gly Glu Gly Asp Tyr Gln Leu Val Gln His  
 170 175 180  
 Glu Ile Leu Cys Ser Met Thr Asn Ser Tyr Glu Val Leu Glu Phe  
 185 190 195  
 Leu Gly Arg Gly Thr Phe Gly Gln Val Ala Lys Cys Trp Lys Arg  
 200 205 210  
 Ser Thr Lys Glu Ile Val Ala Ile Lys Ile Leu Lys Asn His Pro  
 215 220 225  
 Ser Tyr Ala Arg Gln Gly Gln Ile Glu Val Ser Ile Leu Ser Arg  
 230 235 240  
 Leu Ser Ser Glu Asn Ala Asp Glu Tyr Asn Leu Val Arg Ser Tyr  
 245 250 255  
 Glu Cys Phe Gln His Lys Asn His Thr Cys Leu Val Phe Glu Met  
 260 265 270  
 Leu Glu Gln Asn Leu Tyr Asp Phe Leu Lys Gln Asn Lys Phe Ser  
 275 280 285  
 Pro Leu Pro Leu Lys Tyr Ile Arg Pro Ile Leu Gln Gln Val Ala  
 290 295 300  
 Thr Ala Leu Met Lys Leu Lys Ser Leu Gly Leu Ile His Ala Asp  
 305 310 315  
 Leu Lys Pro Glu Asn Ile Met Leu Val Asp Pro Val Arg Gln Pro  
 320 325 330  
 Tyr Arg Val Lys Val Ile Asp Phe Gly Ser Ala Ser His Val Ser  
 335 340 345  
 Lys Ala Val Cys Ser Thr Tyr Leu Gln Ser Arg Tyr Tyr Arg Ala  
 350 355 360  
 Pro Glu Ile Ile Leu Gly Leu Pro Phe Cys Glu Ala Ile Asp Met  
 365 370 375  
 Trp Ser Leu Gly Cys Val Ile Ala Glu Leu Phe Leu Gly Trp Pro  
 380 385 390  
 Leu Tyr Pro Gly Ala Ser Glu Tyr Asp Gln Ile Arg Tyr Ile Ser  
 395 400 405  
 Gln Thr Gln Gly Leu Pro Ala Glu Tyr Leu Leu Ser Ala Gly Thr  
 410 415 420  
 Lys Thr Thr Arg Phe Phe Asn Arg Asp Pro Asn Leu Gly Tyr Pro  
 425 430 435  
 Leu Trp Arg Leu Lys Thr Pro Glu Glu His Glu Leu Glu Thr Gly  
 440 445 450  
 Ile Lys Ser Lys Glu Ala Arg Lys Tyr Ile Phe Asn Cys Leu Asp  
 455 460 465  
 Asp Met Ala Gln Val Asn Met Ser Thr Asp Leu Glu Gly Thr Asp  
 470 475 480  
 Met Leu Ala Glu Lys Ala Asp Arg Arg Glu Tyr Ile Asp Leu Leu  
 485 490 495  
 Lys Lys Met Leu Thr Ile Asp Ala Asp Lys Arg Ile Thr Pro Leu

500	505	510
Lys Thr Leu Asn His Gln Phe Val Thr Met	Thr His Leu Leu Asp	
515	520	525
Phe Pro His Ser Asn His Val Lys Ser Cys	Phe Gln Asn Met Glu	
530	535	540
Ile Cys Lys Arg Arg Val His Met Tyr Asp	Thr Val Ser Gln Ile	
545	550	555
Lys Ser Pro Phe Thr Thr His Val Ala Pro	Asn Thr Ser Thr Asn	
560	565	570
Leu Thr Met Ser Phe Ser Asn Gln Leu Asn	Thr Val His Asn Gln	
575	580	585
Ala Ser Val Leu Ala Ser Ser Ser Thr Ala	Ala Ala Ala Ala Thr	
590	595	600
Ser Leu Ala Asn Ser Asp Val Ser Leu Leu	Asn Tyr Gln Ser Ala	
605	610	615
Leu Tyr Pro Ser Ser Ala Ala Pro Val Pro	Gly Val Ala Gln Gln	
620	625	630
Gly Val Ser Leu Gln Pro Gly Thr Thr Gln	Ile Cys Thr Gln Thr	
635	640	645
Asp Pro Phe Gln Gln Thr Phe Ile Val Cys	Pro Pro Ala Phe Gln	
650	655	660
Thr Gly Leu Gln Ala Thr Thr Lys His Ser	Gly Phe Pro Val Arg	
665	670	675
Met Asp Asn Ala Val Pro Ile Val Pro Gln	Ala Pro Ala Ala Gln	
680	685	690
Pro Leu Gln Ile Gln Ser Gly Val Leu Thr	Gln Gly Ser Cys Thr	
695	700	705
Pro Leu Met Val Ala Thr Leu His Pro Gln	Val Ala Thr Ile Thr	
710	715	720
Pro Gln Tyr Ala Val Pro Phe Thr Leu Ser	Cys Ala Ala Gly Arg	
725	730	735
Pro Ala Leu Val Glu Gln Thr Ala Ala Val	Leu Gln Ala Trp Pro	
740	745	750
Gly Gly Thr Gln Gln Ile Leu Leu Pro Ser	Thr Trp Gln Gln Leu	
755	760	765
Pro Gly Val Ala Leu His Asn Ser Val Gln	Pro Thr Ala Met Ile	
770	775	780
Pro Glu Ala Met Gly Ser Gly Gln Gln Leu	Ala Asp Trp Arg Asn	
785	790	795
Ala His Ser His Gly Asn Gln Tyr Ser Thr	Ile Met Gln Gln Pro	
800	805	810
Ser Leu Leu Thr Asn His Val Thr Leu Ala	Thr Ala Gln Pro Leu	
815	820	825
Asn Val Gly Val Ala His Val Val Arg Gln	Gln Gln Ser Ser Ser	
830	835	840
Leu Pro Ser Lys Lys Asn Lys Gln Ser Ala	Pro Val Ser Ser Lys	
845	850	855
Ser Ser Leu Asp Val Leu Pro Ser Gln Val	Tyr Ser Leu Val Gly	
860	865	870
Ser Ser Pro Leu Arg Thr Thr Ser Ser Tyr	Asn Ser Leu Val Pro	
875	880	885
Val Gln Asp Gln His Gln Pro Ile Ile Ile	Pro Asp Thr Pro Ser	
890	895	900
Pro Pro Val Ser Val Ile Thr Ile Arg Ser	Asp Thr Asp Glu Glu	
905	910	915
Glu Asp Asn Lys Tyr Lys Pro Ser Ser Ser	Gly Leu Lys Pro Arg	
920	925	930
Ser Asn Val Ile Ser Tyr Val Thr Val Asn	Asp Ser Pro Asp Ser	
935	940	945
Asp Ser Ser Leu Ser Ser Pro Tyr Ser Thr	Asp Thr Leu Ser Ala	
950	955	960
Leu Arg Gly Asn Ser Gly Ser Val Leu Glu	Gly Pro Gly Arg Val	
965	970	975

Val Ala Asp Gly Thr Gly Thr Arg Thr Ile Ile Val Pro Pro Leu  
 980 985 990  
 Lys Thr Gln Leu Gly Asp Cys Thr Val Ala Thr Gln Ala Ser Gly  
 995 1000 1005  
 Leu Leu Ser Asn Lys Thr Lys Pro Val Ala Ser Val Ser Gly Gln  
 1010 1015 1020  
 Ser Ser Gly Cys Cys Ile Thr Pro Thr Gly Tyr Arg Ala Gln Arg  
 1025 1030 1035  
 Gly Gly Thr Ser Ala Ala Gln Pro Leu Asn Leu Ser Gln Asn Gln  
 1040 1045 1050  
 Gln Ser Ser Ala Ala Pro Thr Ser Gln Glu Arg Ser Ser Asn Pro  
 1055 1060 1065  
 Ala Pro Arg Arg Gln Gln Ala Phe Val Ala Pro Leu Ser Gln Ala  
 1070 1075 1080  
 Pro Tyr Thr Phe Gln His Gly Ser Pro Leu His Ser Thr Gly His  
 1085 1090 1095  
 Pro His Leu Ala Pro Ala Pro Ala His Leu Pro Ser Gln Ala His  
 1100 1105 1110  
 Leu Tyr Thr Tyr Ala Ala Pro Thr Ser Ala Ala Leu Gly Ser  
 1115 1120 1125  
 Thr Ser Ser Ile Ala His Leu Phe Ser Pro Gln Gly Ser Ser Arg  
 1130 1135 1140  
 His Ala Ala Ala Tyr Thr His Pro Ser Thr Leu Val His Gln  
 1145 1150 1155  
 Val Pro Val Ser Val Gly Pro Ser Leu Leu Thr Ser Ala Ser Val  
 1160 1165 1170  
 Ala Pro Ala Gln Tyr Gln His Gln Phe Ala Thr Gln Ser Tyr Ile  
 1175 1180 1185  
 Gly Ser Ser Arg Gly Ser Thr Ile Tyr Thr Gly Tyr Pro Leu Ser  
 1190 1195 1200  
 Pro Thr Lys Ile Ser Gln Tyr Ser Tyr Leu  
 1205 1210

<210> 2  
 <211> 357  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7472693CD1

<400> 2

Met Ala Arg Glu Asn Gly Glu Ser Ser Ser Trp Lys Lys Gln		
1 5 10	15	
Ala Glu Asp Ile Lys Lys Ile Phe Glu Phe Lys Glu Thr Leu Gly		
20 25	30	
Thr Gly Ala Phe Ser Glu Val Val Leu Ala Glu Glu Lys Ala Thr		
35 40	45	
Gly Lys Leu Phe Ala Val Lys Cys Ile Pro Lys Lys Ala Leu Lys		
50 55	60	
Gly Lys Glu Ser Ser Ile Glu Asn Glu Ile Ala Val Leu Arg Lys		
65 70	75	
Ile Lys His Glu Asn Ile Val Ala Leu Glu Asp Ile Tyr Glu Ser		
80 85	90	
Pro Asn His Leu Tyr Leu Val Met Gln Leu Val Ser Gly Gly Glu		
95 100	105	
Leu Phe Asp Arg Ile Val Glu Lys Gly Phe Tyr Thr Glu Lys Asp		
110 115	120	
Ala Ser Thr Leu Ile Arg Gln Val Leu Asp Ala Val Tyr Tyr Leu		
125 130	135	
His Arg Met Gly Ile Val His Arg Asp Leu Lys Pro Glu Asn Leu		
140 145	150	

Leu Tyr Tyr Ser Gln Asp Glu Glu Ser Lys Ile Met Ile Ser Asp  
     155                   160                   165  
 Phe Gly Leu Ser Lys Met Glu Gly Lys Gly Asp Val Met Ser Thr  
     170                   175                   180  
 Ala Cys Gly Thr Pro Gly Tyr Val Ala Pro Glu Val Leu Ala Gln  
     185                   190                   195  
 Lys Pro Tyr Ser Lys Ala Val Asp Cys Trp Ser Ile Gly Val Ile  
     200                   205                   210  
 Ala Tyr Ile Leu Leu Cys Gly Tyr Pro Pro Phe Tyr Asp Glu Asn  
     215                   220                   225  
 Asp Ser Lys Leu Phe Glu Gln Ile Leu Lys Ala Glu Tyr Glu Phe  
     230                   235                   240  
 Asp Ser Pro Tyr Trp Asp Asp Ile Ser Asp Ser Ala Lys Asp Phe  
     245                   250                   255  
 Ile Arg Asn Leu Met Glu Lys Asp Pro Asn Lys Arg Tyr Thr Cys  
     260                   265                   270  
 Glu Gln Ala Ala Arg His Pro Trp Ile Ala Gly Asp Thr Ala Leu  
     275                   280                   285  
 Asn Lys Asn Ile His Glu Ser Val Ser Ala Gln Ile Arg Lys Asn  
     290                   295                   300  
 Phe Ala Lys Ser Lys Trp Arg Gln Ala Phe Asn Ala Thr Ala Val  
     305                   310                   315  
 Val Arg His Met Arg Lys Leu His Leu Gly Ser Ser Leu Asp Ser  
     320                   325                   330  
 Ser Asn Ala Ser Val Ser Ser Ser Leu Ser Leu Ala Ser Gln Lys  
     335                   340                   345  
 Asp Cys Ala Tyr Val Ala Lys Pro Glu Ser Leu Ser  
     350                   355

<210> 3  
 <211> 419  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3107952CD1

<400> 3

Met Ala Gly Ser Gly Cys Ala Trp Gly Ala Glu Pro Pro Arg Phe  
     1                   5                   10                   15  
 Leu Glu Ala Phe Gly Arg Leu Trp Gln Val Gln Ser Arg Leu Gly  
     20                   25                   30  
 Ser Gly Ser Ser Ala Ser Val Tyr Arg Val Arg Cys Cys Gly Asn  
     35                   40                   45  
 Pro Gly Ser Pro Gly Ala Leu Lys Gln Phe Leu Pro Pro Gly  
     50                   55                   60  
 Thr Thr Gly Ala Ala Ala Ser Ala Ala Glu Tyr Gly Phe Arg Lys  
     65                   70                   75  
 Glu Arg Ala Ala Leu Glu Gln Leu Gln Gly His Arg Asn Ile Val  
     80                   85                   90  
 Thr Leu Tyr Gly Val Phe Thr Ile His Phe Ser Pro Asn Val Pro  
     95                   100                   105  
 Ser Arg Cys Leu Leu Glu Leu Leu Asp Val Ser Val Ser Glu  
     110                   115                   120  
 Leu Leu Leu Tyr Ser Ser His Gln Gly Cys Ser Met Trp Met Ile  
     125                   130                   135  
 Gln His Cys Ala Arg Asp Val Leu Glu Ala Leu Ala Phe Leu His  
     140                   145                   150  
 His Glu Gly Tyr Val His Ala Asp Leu Lys Pro Arg Asn Ile Leu  
     155                   160                   165  
 Trp Ser Ala Glu Asn Glu Cys Phe Lys Leu Ile Asp Phe Gly Leu  
     170                   175                   180

Ser Phe Lys Glu Gly Asn Gln Asp Val Lys Tyr Ile Gln Thr Asp  
 185 190 195  
 Gly Tyr Arg Ala Pro Glu Ala Glu Leu Gln Asn Cys Leu Ala Gln  
 200 205 210  
 Ala Gly Leu Gln Ser Asp Thr Glu Cys Thr Ser Ala Val Asp Leu  
 215 220 225  
 Trp Ser Leu Gly Ile Ile Leu Leu Glu Met Phe Ser Gly Met Lys  
 230 235 240  
 Leu Lys His Thr Val Arg Ser Gln Glu Trp Lys Ala Asn Ser Ser  
 245 250 255  
 Ala Ile Ile Asp His Ile Phe Ala Ser Lys Ala Val Val Asn Ala  
 260 265 270  
 Ala Ile Pro Ala Tyr His Leu Arg Asp Leu Ile Lys Ser Met Leu  
 275 280 285  
 His Asp Asp Pro Ser Arg Arg Ile Pro Ala Glu Met Ala Leu Cys  
 290 295 300  
 Ser Pro Phe Phe Ser Ile Pro Phe Ala Pro His Ile Glu Asp Leu  
 305 310 315  
 Val Met Leu Pro Thr Pro Val Leu Arg Leu Leu Asn Val Leu Asp  
 320 325 330  
 Asp Asp Tyr Leu Glu Asn Glu Glu Glu Tyr Glu Asp Val Val Glu  
 335 340 345  
 Asp Val Lys Glu Glu Cys Gln Lys Tyr Gly Pro Val Val Ser Leu  
 350 355 360  
 Leu Val Pro Lys Gly Asn Pro Gly Arg Gly Gln Val Phe Val Glu  
 365 370 375  
 Tyr Ala Asn Ala Gly Asp Ser Lys Ala Ala Gln Lys Leu Leu Thr  
 380 385 390  
 Gly Arg Met Phe Asp Gly Lys Phe Val Val Ala Thr Phe Tyr Pro  
 395 400 405  
 Leu Ser Ala Tyr Lys Arg Gly Tyr Leu Tyr Gln Thr Leu Leu  
 410 415

<210> 4  
 <211> 624  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5544420CD1

<400> 4

Met Asn Arg Tyr Thr	Thr Met Arg Gln Leu Gly Asp Gly Thr Tyr	
1	5 10	15
Gly Ser Val Leu Met	Gly Lys Ser Asn Glu Ser Gly Glu Leu Val	
	20 25	30
Ala Ile Lys Arg Met	Lys Arg Lys Phe Tyr Ser Trp Asp Glu Cys	
	35 40	45
Met Asn Leu Arg Glu	Val Lys Ser Leu Lys Lys Leu Asn His Ala	
	50 55	60
Asn Val Ile Lys Leu	Lys Glu Val Ile Arg Glu Asn Asp His Leu	
	65 70	75
Tyr Phe Ile Phe Glu	Tyr Met Lys Glu Asn Leu Tyr Gln Leu Met	
	80 85	90
Lys Asp Arg Asn Lys	Leu Phe Pro Glu Ser Val Ile Arg Asn Ile	
	95 100	105
Met Tyr Gln Ile Leu	Gln Gly Leu Ala Phe Ile His Lys His Gly	
	110 115	120
Phe Phe His Arg Asp	Met Lys Pro Glu Asn Leu Leu Cys Met Gly	
	125 130	135
Pro Glu Leu Val Lys	Ile Ala Asp Phe Gly Leu Ala Arg Glu Leu	
	140 145	150

Arg Ser Gln Pro Pro Tyr Thr Asp Tyr Val Ser Thr Arg Trp Tyr		
155	160	165
Arg Ala Pro Glu Val Leu Leu Arg Ser Ser Val Tyr Ser Ser Pro		
170	175	180
Ile Asp Val Trp Ala Val Gly Ser Ile Met Ala Glu Leu Tyr Met		
185	190	195
Leu Arg Pro Leu Phe Pro Gly Thr Ser Glu Val Asp Glu Ile Phe		
200	205	210
Lys Ile Cys Gln Val Leu Gly Thr Pro Lys Lys Ser Asp Trp Pro		
215	220	225
Glu Gly Tyr Gln Leu Ala Ser Ser Met Asn Phe Arg Phe Pro Gln		
230	235	240
Cys Val Pro Ile Asn Leu Lys Thr Leu Ile Pro Asn Ala Ser Asn		
245	250	255
Glu Ala Ile Gln Leu Met Thr Glu Met Leu Asn Trp Asp Pro Lys		
260	265	270
Lys Arg Pro Thr Ala Ser Gln Ala Leu Lys His Pro Tyr Phe Gln		
275	280	285
Val Gly Gln Val Leu Gly Pro Ser Ser Asn His Leu Glu Ser Lys		
290	295	300
Gln Ser Leu Asn Lys Gln Leu Gln Pro Leu Glu Ser Lys Pro Ser		
305	310	315
Leu Val Glu Val Glu Pro Lys Pro Leu Pro Asp Ile Ile Asp Gln		
320	325	330
Val Val Gly Gln Pro Gln Pro Lys Thr Ser Gln Gln Pro Leu Gln		
335	340	345
Pro Ile Gln Pro Pro Gln Asn Leu Ser Val Gln Gln Pro Pro Lys		
350	355	360
Gln Gln Ser Gln Glu Lys Pro Pro Gln Thr Leu Phe Pro Ser Ile		
365	370	375
Val Lys Asn Met Pro Thr Lys Pro Asn Gly Thr Leu Ser His Lys		
380	385	390
Ser Gly Arg Arg Arg Trp Gly Gln Thr Ile Phe Lys Ser Gly Asp		
395	400	405
Ser Trp Glu Glu Leu Glu Asp Tyr Asp Phe Gly Ala Ser His Ser		
410	415	420
Lys Lys Pro Ser Met Gly Val Phe Lys Glu Lys Arg Lys Lys Asp		
425	430	435
Ser Pro Phe Arg Leu Pro Glu Pro Val Pro Ser Gly Ser Asn His		
440	445	450
Ser Thr Gly Glu Asn Lys Ser Leu Pro Ala Val Thr Ser Leu Lys		
455	460	465
Ser Asp Ser Glu Leu Ser Thr Ala Pro Thr Ser Lys Gln Tyr Tyr		
470	475	480
Leu Lys Gln Ser Arg Tyr Leu Pro Gly Val Asn Pro Lys Lys Val		
485	490	495
Ser Leu Ile Ala Ser Gly Lys Glu Ile Asn Pro His Thr Trp Ser		
500	505	510
Asn Gln Leu Phe Pro Lys Ser Leu Gly Pro Val Gly Ala Glu Leu		
515	520	525
Ala Phe Lys Arg Ser Asn Ala Glu Glu Lys Leu Gly Ser Tyr Ala		
530	535	540
Thr Tyr Asn Gln Ser Gly Tyr Ile Pro Ser Phe Leu Lys Lys Glu		
545	550	555
Val Gln Ser Ala Gly Gln Arg Ile His Leu Ala Pro Leu Asn Ala		
560	565	570
Thr Ala Ser Glu Tyr Thr Trp Asn Thr Lys Thr Gly Arg Gly Gln		
575	580	585
Phe Ser Gly Arg Thr Tyr Asn Pro Thr Ala Lys Asn Leu Asn Ile		
590	595	600
Val Asn Arg Ala Gln Pro Ile Pro Ser Val His Gly Arg Thr Asp		
605	610	615
Trp Val Ala Lys Tyr Gly His Arg		

620

<210> 5  
 <211> 878  
 <212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7472832CD1

<400> 5

Met Ala Thr Ala Pro Ser Tyr Pro Ala Gly Leu Pro Gly Ser Pro  
 1 5 10 15  
 Gly Pro Gly Ser Pro Pro Pro Gly Gly Leu Glu Leu Gln Ser  
 20 25 30  
 Pro Pro Pro Leu Leu Pro Gln Ile Pro Ala Pro Gly Ser Gly Val  
 35 40 45  
 Ser Phe His Ile Gln Ile Gly Leu Thr Arg Glu Phe Val Leu Leu  
 50 55 60  
 Pro Ala Ala Ser Glu Leu Ala His Val Lys Gln Leu Ala Cys Ser  
 65 70 75  
 Ile Val Asp Gln Lys Phe Pro Glu Cys Gly Phe Tyr Gly Leu Tyr  
 80 85 90  
 Asp Lys Ile Leu Leu Phe Lys His Asp Pro Thr Ser Ala Asn Leu  
 95 100 105  
 Leu Gln Leu Val Arg Ser Ser Gly Asp Ile Gln Glu Gly Asp Leu  
 110 115 120  
 Val Glu Val Val Leu Ser Ala Ser Ala Thr Phe Glu Asp Phe Gln  
 125 130 135  
 Ile Arg Pro His Ala Leu Thr Val His Ser Tyr Arg Ala Pro Ala  
 140 145 150  
 Phe Cys Asp His Cys Gly Glu Met Leu Phe Gly Leu Val Arg Gln  
 155 160 165  
 Gly Leu Lys Cys Asp Gly Cys Gly Leu Asn Tyr His Lys Arg Cys  
 170 175 180  
 Ala Phe Ser Ile Pro Asn Asn Cys Ser Gly Ala Arg Lys Arg Arg  
 185 190 195  
 Leu Ser Ser Thr Ser Leu Ala Ser Gly His Ser Val Arg Leu Gly  
 200 205 210  
 Thr Ser Glu Ser Leu Pro Cys Thr Ala Glu Glu Leu Ser Arg Ser  
 215 220 225  
 Thr Thr Glu Leu Leu Pro Arg Arg Pro Pro Ser Ser Ser Ser  
 230 235 240  
 Ser Ser Ala Ser Ser Tyr Thr Gly Arg Pro Ile Glu Leu Asp Lys  
 245 250 255  
 Met Leu Leu Ser Lys Val Lys Val Pro His Thr Phe Leu Ile His  
 260 265 270  
 Ser Tyr Thr Arg Pro Thr Val Cys Gln Ala Cys Lys Lys Leu  
 275 280 285  
 Lys Gly Leu Phe Arg Gln Gly Leu Gln Cys Lys Asp Cys Lys Phe  
 290 295 300  
 Asn Cys His Lys Arg Cys Ala Thr Arg Val Pro Asn Asp Cys Leu  
 305 310 315  
 Gly Glu Ala Leu Ile Asn Gly Asp Val Pro Met Glu Glu Ala Thr  
 320 325 330  
 Asp Phe Ser Glu Ala Asp Lys Ser Ala Leu Met Asp Glu Ser Glu  
 335 340 345  
 Asp Ser Gly Val Ile Pro Gly Ser His Ser Glu Asn Ala Leu His  
 350 355 360  
 Ala Ser Glu Glu Glu Gly Glu Gly Lys Ala Gln Ser Ser  
 365 370 375  
 Leu Gly Tyr Ile Pro Leu Met Arg Val Val Gln Ser Val Arg His

380	385	390
Thr Thr Arg Lys Ser Ser Thr Thr Leu Arg Glu Gly Trp Val Val		
395	400	405
His Tyr Ser Asn Lys Asp Thr Leu Arg Lys Arg His Tyr Trp Arg		
410	415	420
Leu Asp Cys Lys Cys Ile Thr Leu Phe Gln Asn Asn Thr Thr Asn		
425	430	435
Arg Tyr Tyr Lys Glu Ile Pro Leu Ser Glu Ile Leu Thr Val Glu		
440	445	450
Ser Ala Gln Asn Phe Ser Ile Val Pro Pro Gly Thr Asn Pro His		
455	460	465
Cys Phe Glu Ile Val Thr Ala Asn Ala Thr Tyr Phe Val Gly Glu		
470	475	480
Met Pro Gly Gly Thr Pro Gly Gly Pro Ser Gly Gln Gly Ala Glu		
485	490	495
Ala Ala Arg Gly Trp Glu Thr Ala Ile Arg Gln Ala Leu Met Pro		
500	505	510
Val Ile Leu Gln Asp Ala Pro Ser Ala Pro Gly His Ala Pro His		
515	520	525
Arg Gln Ala Ser Leu Ser Ile Ser Val Ser Asn Ser Gln Ile Gln		
530	535	540
Glu Asn Val Asp Ile Ala Thr Val Tyr Gln Ile Phe Pro Asp Glu		
545	550	555
Val Leu Gly Ser Gly Gln Phe Gly Val Val Tyr Gly Gly Lys His		
560	565	570
Arg Lys Thr Gly Arg Asp Val Ala Val Lys Val Ile Asp Lys Leu		
575	580	585
Arg Phe Pro Thr Lys Gln Glu Ser Gln Leu Arg Asn Glu Val Ala		
590	595	600
Ile Leu Gln Ser Leu Arg His Pro Gly Ile Val Asn Leu Glu Cys		
605	610	615
Met Phe Glu Thr Pro Glu Lys Val Phe Val Val Met Glu Lys Leu		
620	625	630
His Gly Asp Met Leu Glu Met Ile Leu Ser Ser Glu Lys Gly Arg		
635	640	645
Leu Pro Glu Arg Leu Thr Lys Phe Leu Ile Thr Gln Ile Leu Val		
650	655	660
Ala Leu Arg His Leu His Phe Lys Asn Ile Val His Cys Asp Leu		
665	670	675
Lys Pro Glu Asn Val Leu Leu Ala Ser Ala Asp Pro Phe Pro Gln		
680	685	690
Val Lys Leu Cys Asp Phe Gly Phe Ala Arg Ile Ile Gly Glu Lys		
695	700	705
Ser Phe Arg Arg Ser Val Val Gly Thr Pro Ala Tyr Leu Ala Pro		
710	715	720
Glu Val Leu Leu Asn Gln Gly Tyr Asn Arg Ser Leu Asp Met Trp		
725	730	735
Ser Val Gly Val Ile Met Tyr Val Ser Leu Ser Gly Thr Phe Pro		
740	745	750
Phe Asn Glu Asp Glu Asp Ile Asn Asp Gln Ile Gln Asn Ala Ala		
755	760	765
Phe Met Tyr Pro Ala Ser Pro Trp Ser His Ile Ser Ala Gly Ala		
770	775	780
Ile Asp Leu Ile Asn Asn Leu Leu Gln Val Lys Met Arg Lys Arg		
785	790	795
Tyr Ser Val Asp Lys Ser Leu Ser His Pro Trp Leu Gln Glu Tyr		
800	805	810
Gln Thr Trp Leu Asp Leu Arg Glu Leu Glu Gly Lys Met Gly Glu		
815	820	825
Arg Tyr Ile Thr His Glu Ser Asp Asp Ala Arg Trp Glu Gln Phe		
830	835	840
Ala Ala Glu His Pro Leu Pro Gly Ser Gly Leu Pro Thr Asp Arg		
845	850	855

Asp Leu Gly Gly Ala Cys Pro Pro Gln Asp His Asp Met Gln Gly  
 860 865 870  
 Leu Ala Glu Arg Ile Ser Val Leu  
 875

<210> 6  
 <211> 440  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1551456CD1

<400> 6

Met	Ser	Lys	Leu	Arg	Met	Lys	Arg	Arg	Ala	Ser	Asp	Arg	Gly	Ala
1					5				10				15	
Gly	Glu	Thr	Ser	Ala	Arg	Ala	Lys	Ala	Leu	Gly	Ser	Gly	Ile	Ser
					20				25				30	
Gly	Asn	Asn	Ala	Lys	Arg	Ala	Gly	Pro	Phe	Ile	Leu	Gly	Pro	Arg
					35				40				45	
Leu	Gly	Asn	Ser	Pro	Val	Pro	Ser	Ile	Val	Gln	Cys	Leu	Ala	Arg
					50				55				60	
Lys	Asp	Gly	Thr	Asp	Asp	Phe	Tyr	Gln	Leu	Lys	Ile	Leu	Thr	Leu
					65				70				75	
Glu	Glu	Arg	Gly	Asp	Gln	Gly	Ile	Glu	Ser	Gln	Glu	Glu	Arg	Gln
					80				85				90	
Gly	Lys	Met	Leu	Leu	His	Thr	Glu	Tyr	Ser	Leu	Leu	Ser	Leu	Leu
					95				100				105	
His	Thr	Gln	Asp	Gly	Val	Val	His	His	His	Gly	Leu	Phe	Gln	Asp
					110				115				120	
Arg	Thr	Cys	Glu	Ile	Val	Glu	Asp	Thr	Glu	Ser	Ser	Arg	Met	Val
					125				130				135	
Lys	Lys	Met	Lys	Lys	Arg	Ile	Cys	Leu	Val	Leu	Asp	Cys	Leu	Cys
					140				145				150	
Ala	His	Asp	Phe	Ser	Asp	Lys	Thr	Ala	Asp	Leu	Ile	Asn	Leu	Gln
					155				160				165	
His	Tyr	Val	Ile	Lys	Glu	Lys	Arg	Leu	Ser	Glu	Arg	Glu	Thr	Val
					170				175				180	
Val	Ile	Phe	Tyr	Asp	Val	Val	Arg	Val	Val	Glu	Ala	Leu	His	Gln
					185				190				195	
Lys	Asn	Ile	Val	His	Arg	Asp	Leu	Lys	Leu	Gly	Asn	Met	Val	Leu
					200				205				210	
Asn	Lys	Arg	Thr	His	Arg	Ile	Thr	Ile	Thr	Asn	Phe	Cys	Leu	Gly
					215				220				225	
Lys	His	Leu	Val	Ser	Glu	Gly	Asp	Leu	Leu	Lys	Asp	Gln	Arg	Gly
					230				235				240	
Ser	Pro	Ala	Tyr	Ile	Ser	Pro	Asp	Val	Leu	Ser	Gly	Arg	Pro	Tyr
					245				250				255	
Arg	Gly	Lys	Pro	Ser	Asp	Met	Trp	Ala	Leu	Gly	Val	Val	Leu	Phe
					260				265				270	
Thr	Met	Leu	Tyr	Gly	Gln	Phe	Pro	Phe	Tyr	Asp	Ser	Ile	Pro	Gln
					275				280				285	
Glu	Leu	Phe	Arg	Lys	Ile	Lys	Ala	Ala	Glu	Tyr	Thr	Ile	Pro	Glu
					290				295				300	
Asp	Gly	Arg	Val	Ser	Glu	Asn	Thr	Val	Cys	Leu	Ile	Arg	Lys	Leu
					305				310				315	
Leu	Val	Leu	Asp	Pro	Gln	Gln	Arg	Leu	Ala	Ala	Ala	Asp	Val	Leu
					320				325				330	
Glu	Ala	Leu	Ser	Ala	Ile	Ile	Ala	Ser	Trp	Gln	Ser	Leu	Ser	Ser
					335				340				345	
Leu	Ser	Gly	Pro	Leu	Gln	Val	Val	Pro	Asp	Ile	Asp	Asp	Gln	Met
					350				355				360	

Ser Asn Ala Asp Ser Ser Gln Glu Ala Lys Val Thr Glu Glu Cys  
 365 370 375  
 Ser Gln Tyr Glu Phe Glu Asn Tyr Met Arg Gln Gln Leu Leu Leu  
 380 385 390  
 Ala Glu Glu Lys Ser Ser Ile His Asp Ala Arg Ser Trp Val Pro  
 395 400 405  
 Lys Arg Gln Phe Gly Ser Ala Pro Pro Val Arg Arg Leu Gly His  
 410 415 420  
 Asp Ala Gln Pro Met Thr Ser Leu Asp Thr Ala Ile Leu Ala Gln  
 425 430 435  
 Arg Tyr Leu Arg Lys  
 440

<210> 7  
 <211> 923  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2589355CD1

<400> 7

Met Ala Arg Gly Thr Cys Ser Ala Gly Arg Ser Gly Trp Gly Ser  
 1 5 10 15  
 Thr Thr Ser Arg Ala Arg Trp Ala Ser Gly Asn Phe Ala Val Val  
 20 25 30  
 Lys Leu Gly Arg His Arg Ile Thr Lys Thr Glu Val Ala Ile Lys  
 35 40 45  
 Ile Ile Asp Lys Ser Gln Pro Trp Met His Val Asn Leu Glu Lys  
 50 55 60  
 Ile Tyr Arg Glu Val Gln Ile Met Lys Met Leu Asp His Pro His  
 65 70 75  
 Ile Ile Lys Leu Tyr Gln Val Met Glu Thr Lys Ser Met Leu Tyr  
 80 85 90  
 Leu Val Thr Glu Tyr Ala Lys Asn Gly Glu Ile Phe Asp Tyr Leu  
 95 100 105  
 Ala Asn His Gly Arg Leu Asn Glu Ser Glu Ala Arg Arg Lys Phe  
 110 115 120  
 Trp Gln Ile Leu Ser Ala Val Asp Tyr Cys His Gly Arg Lys Ile  
 125 130 135  
 Val His Arg Asp Leu Lys Ala Glu Asn Leu Leu Leu Asp Asn Asn  
 140 145 150  
 Met Asn Ile Lys Ile Ala Asp Phe Gly Phe Gly Asn Phe Phe Lys  
 155 160 165  
 Ser Gly Glu Leu Leu Ala Thr Trp Cys Gly Ser Pro Pro Tyr Ala  
 170 175 180  
 Ala Pro Glu Val Phe Glu Gly Gln Gln Tyr Glu Gly Pro Gln Leu  
 185 190 195  
 Asp Ile Trp Ser Met Gly Val Val Leu Tyr Val Leu Val Cys Gly  
 200 205 210  
 Ala Leu Pro Phe Asp Gly Pro Thr Leu Pro Ile Leu Arg Gln Arg  
 215 220 225  
 Val Leu Glu Gly Arg Phe Arg Ile Pro Tyr Phe Met Ser Glu Asp  
 230 235 240  
 Cys Glu His Leu Ile Arg Arg Met Leu Val Leu Asp Pro Ser Lys  
 245 250 255  
 Arg Leu Thr Ile Ala Gln Ile Lys Glu His Lys Trp Met Leu Ile  
 260 265 270  
 Glu Val Pro Val Gln Arg Pro Val Leu Tyr Pro Gln Glu Gln Glu  
 275 280 285  
 Asn Glu Pro Ser Ile Gly Glu Phe Asn Glu Gln Val Leu Arg Leu  
 290 295 300

Met His Ser Leu Gly Ile Asp Gln Gln Lys Thr Ile Glu Ser Leu  
 305 310 315  
 Gln Asn Lys Ser Tyr Asn His Phe Ala Ala Ile Tyr Phe Leu Leu  
 320 325 330  
 Val Glu Arg Leu Lys Ser His Arg Ser Ser Phe Pro Val Glu Gln  
 335 340 345  
 Arg Leu Asp Gly Arg Gln Arg Arg Pro Ser Thr Ile Ala Glu Gln  
 350 355 360  
 Thr Val Ala Lys Ala Gln Thr Val Gly Leu Pro Val Thr Met His  
 365 370 375  
 Ser Pro Asn Met Arg Leu Leu Arg Ser Ala Leu Leu Pro Gln Ala  
 380 385 390  
 Ser Asn Val Glu Ala Phe Ser Phe Pro Ala Ser Gly Cys Gln Ala  
 395 400 405  
 Glu Ala Ala Phe Met Glu Glu Glu Cys Val Asp Thr Pro Lys Val  
 410 415 420  
 Asn Gly Cys Leu Leu Asp Pro Val Pro Pro Val Leu Val Arg Lys  
 425 430 435  
 Gly Cys Gln Ser Leu Pro Ser Asn Met Met Glu Thr Ser Ile Asp  
 440 445 450  
 Glu Gly Leu Glu Thr Glu Gly Glu Ala Glu Glu Asp Pro Ala His  
 455 460 465  
 Ala Phe Glu Ala Phe Gln Ser Thr Arg Ser Gly Gln Arg Arg His  
 470 475 480  
 Thr Leu Ser Glu Val Thr Asn Gln Leu Val Val Met Pro Gly Ala  
 485 490 495  
 Gly Lys Ile Phe Ser Met Asn Asp Ser Pro Ser Leu Asp Ser Val  
 500 505 510  
 Asp Ser Glu Tyr Asp Met Gly Ser Val Gln Arg Asp Leu Asn Phe  
 515 520 525  
 Leu Glu Asp Asn Pro Ser Leu Lys Asp Ile Met Leu Ala Asn Gln  
 530 535 540  
 Pro Ser Pro Arg Met Thr Ser Pro Phe Ile Ser Leu Arg Pro Thr  
 545 550 555  
 Asn Pro Ala Met Gln Ala Leu Ser Ser Gln Lys Arg Glu Val His  
 560 565 570  
 Asn Arg Ser Pro Val Ser Phe Arg Glu Gly Arg Arg Ala Ser Asp  
 575 580 585  
 Thr Ser Leu Thr Gln Gly Ile Val Ala Phe Arg Gln His Leu Gln  
 590 595 600  
 Asn Leu Ala Arg Thr Lys Gly Ile Leu Glu Leu Asn Lys Val Gln  
 605 610 615  
 Leu Leu Tyr Glu Gln Ile Gly Pro Glu Ala Asp Pro Asn Leu Ala  
 620 625 630  
 Pro Ala Ala Pro Gln Leu Gln Asp Leu Ala Ser Ser Cys Pro Gln  
 635 640 645  
 Glu Glu Val Ser Gln Gln Glu Ser Val Ser Thr Leu Pro Ala  
 650 655 660  
 Ser Val His Pro Gln Leu Ser Pro Arg Gln Ser Leu Glu Thr Gln  
 665 670 675  
 Tyr Leu Gln His Arg Leu Gln Lys Pro Ser Leu Leu Ser Lys Ala  
 680 685 690  
 Gln Asn Thr Cys Gln Leu Tyr Cys Lys Glu Pro Pro Arg Ser Leu  
 695 700 705  
 Glu Gln Gln Leu Gln Glu His Arg Leu Gln Gln Lys Arg Leu Phe  
 710 715 720  
 Leu Gln Lys Gln Ser Gln Leu Gln Ala Tyr Phe Asn Gln Met Gln  
 725 730 735  
 Ile Ala Glu Ser Ser Tyr Pro Gln Pro Ser Gln Gln Leu Pro Leu  
 740 745 750  
 Pro Arg Gln Glu Thr Pro Pro Pro Ser Gln Gln Ala Pro Pro Phe  
 755 760 765  
 Ser Leu Thr Gln Pro Leu Ser Pro Val Leu Glu Pro Ser Ser Glu

770	775	780
Gln Met Gln Tyr Ser Pro Phe Leu Ser Gln Tyr Gln Glu Met Gln		
785	790	795
Leu Gln Pro Leu Pro Ser Thr Ser Gly Pro Arg Ala Ala Pro Pro		
800	805	810
Leu Pro Thr Gln Leu Gln Gln Gln Pro Pro Pro Pro Pro Pro		
815	820	825
Pro Pro Pro Arg Gln Pro Gly Ala Ala Pro Ala Pro Leu Gln		
830	835	840
Phe Ser Tyr Gln Thr Cys Glu Leu Pro Ser Ala Ala Ser Pro Ala		
845	850	855
Pro Asp Tyr Pro Thr Pro Cys Gln Tyr Pro Val Asp Gly Ala Gln		
860	865	870
Gln Ser Asp Leu Thr Gly Pro Asp Cys Pro Arg Ser Pro Gly Leu		
875	880	885
Gln Glu Ala Pro Ser Ser Tyr Asp Pro Leu Ala Leu Ser Glu Leu		
890	895	900
Pro Gly Leu Phe Asp Cys Glu Met Leu Asp Ala Val Asp Pro Gln		
905	910	915
His Asn Gly Tyr Val Leu Val Asn		
920		

&lt;210&gt; 8

&lt;211&gt; 442

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4357117CD1

&lt;400&gt; 8

Met Arg Ile Val Cys Leu Val Lys Asn Gln Gln Pro Leu Gly Ala			
1	5	10	15
Thr Ile Lys Arg His Glu Met Thr Gly Asp Ile Leu Val Ala Arg			
20	25	30	
Ile Ile His Gly Gly Leu Ala Glu Arg Ser Gly Leu Leu Tyr Ala			
35	40	45	
Gly Asp Lys Leu Val Glu Val Asn Gly Val Ser Val Glu Gly Leu			
50	55	60	
Asp Pro Glu Gln Val Ile His Ile Leu Ala Met Ser Arg Gly Thr			
65	70	75	
Ile Met Phe Lys Val Val Pro Val Ser Asp Pro Pro Val Asn Ser			
80	85	90	
Gln Gln Met Val Tyr Val Arg Ala Met Thr Glu Tyr Trp Pro Gln			
95	100	105	
Glu Asp Pro Asp Ile Pro Cys Met Asp Ala Gly Leu Pro Phe Gln			
110	115	120	
Lys Gly Asp Ile Leu Gln Ile Val Asp Gln Asn Asp Ala Leu Trp			
125	130	135	
Trp Gln Ala Arg Lys Ile Ser Asp Pro Ala Thr Cys Ala Gly Leu			
140	145	150	
Val Pro Ser Asn His Leu Leu Lys Arg Lys Gln Arg Glu Phe Trp			
155	160	165	
Trp Ser Gln Pro Tyr Gln Pro His Thr Cys Leu Lys Ser Thr Ser			
170	175	180	
Asp Lys Glu Glu Phe Val Gly Tyr Gly Gln Lys Phe Phe Ile Gly			
185	190	195	
Arg Phe Ser Pro Leu His Ala Ser Val Cys Cys Thr Gly Ser Cys			
200	205	210	
Tyr Ser Ala Val Gly Ala Pro Tyr Glu Glu Val Val Arg Tyr Gln			
215	220	225	
Arg Arg Pro Ser Asp Lys Tyr Arg Leu Ile Val Leu Ile Gly Pro			

Ser	Gly	Val	Gly	Val	Asn	Glu	Leu	Arg	Arg	Gln	Leu	Ile	Glu	Phe
				245					250					255
Asn	Pro	Ser	His	Phe	Gln	Ser	Ala	Val	Pro	His	Thr	Thr	Arg	Thr
				260					265					270
Lys	Lys	Ser	Tyr	Glu	Met	Asn	Gly	Arg	Glu	Tyr	His	Tyr	Val	Ser
				275					280					285
Lys	Glu	Thr	Phe	Glu	Asn	Leu	Ile	Tyr	Ser	His	Arg	Met	Leu	Glu
				290					295					300
Tyr	Gly	Glu	Tyr	Lys	Gly	His	Leu	Tyr	Gly	Thr	Ser	Val	Asp	Ala
				305					310					315
Val	Gln	Thr	Val	Leu	Val	Glu	Gly	Lys	Ile	Cys	Val	Met	Asp	Leu
				320					325					330
Glu	Pro	Gln	Asp	Ile	Gln	Gly	Val	Arg	Thr	His	Glu	Leu	Lys	Pro
				335					340					345
Tyr	Val	Ile	Phe	Ile	Lys	Pro	Ser	Asn	Met	Arg	Cys	Met	Lys	Gln
				350					355					360
Ser	Arg	Lys	Asn	Ala	Lys	Val	Ile	Thr	Asp	Tyr	Tyr	Val	Asp	Met
				365					370					375
Lys	Phe	Lys	Asp	Glu	Asp	Leu	Gln	Glu	Met	Glu	Asn	Leu	Ala	Gln
				380					385					390
Arg	Met	Glu	Thr	Gln	Phe	Gly	Gln	Phe	Phe	Asp	His	Val	Ile	Val
				395					400					405
Asn	Asp	Ser	Leu	His	Asp	Ala	Cys	Ala	Gln	Leu	Leu	Ser	Ala	Ile
				410					415					420
Gln	Lys	Ala	Gln	Glu	Glu	Pro	Gln	Trp	Val	Pro	Ala	Thr	Trp	Ile
				425					430					435
Ser	Ser	Asp	Thr	Glu	Ser	Gln								
				440										

<210> 9  
 <211> 1046  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5511992CD1

Met	Glu	Pro	Ser	Arg	Ala	Leu	Leu	Gly	Cys	Leu	Ala	Ser	Ala	Ala
1				5					10					15
Ala	Ala	Ala	Pro	Pro	Gly	Glu	Asp	Gly	Ala	Gly	Ala	Gly	Ala	Glu
				20					25					30
Glu	Ala	Ala	Ala	Ala	Ala	Val	Gly							
				35					40					45
Gly	Glu	Leu	Gly	Cys	Asp	Ala	Pro	Leu	Pro	Tyr	Trp	Thr	Ala	Val
				50					55					60
Phe	Glu	Tyr	Glu	Ala	Ala	Gly	Glu	Asp	Glu	Leu	Thr	Leu	Arg	Leu
				65					70					75
Gly	Asp	Val	Val	Glu	Val	Leu	Ser	Lys	Asp	Ser	Gln	Val	Ser	Gly
				80					85					90
Asp	Glu	Gly	Trp	Trp	Thr	Gly	Gln	Leu	Asn	Gln	Arg	Val	Gly	Ile
				95					100					105
Phe	Pro	Ser	Asn	Tyr	Val	Thr	Pro	Arg	Ser	Ala	Phe	Ser	Ser	Arg
				110					115					120
Cys	Gln	Pro	Gly	Gly	Glu	Glu	Glu	Ile	Asp	Phe	Ala	Glu	Leu	Thr
				125					130					135
Leu	Glu	Glu	Ile	Ile	Gly	Ile	Gly	Gly	Phe	Gly	Lys	Val	Tyr	Arg
				140					145					150
Ala	Phe	Trp	Ile	Gly	Asp	Glu	Val	Ala	Val	Lys	Ala	Ala	Arg	His
				155					160					165
Asp	Pro	Asp	Glu	Asp	Ile	Ser	Gln	Thr	Ile	Glu	Asn	Val	Arg	Gln

	170	175	180
Glu Ala Lys Leu Phe Ala Met Leu Lys His Pro Asn Ile Ile Ala			
185	190	195	
Leu Arg Gly Val Cys Leu Lys Glu Pro Asn Leu Cys Leu Val Met			
200	205	210	
Glu Phe Ala Arg Gly Gly Pro Leu Asn Arg Val Leu Ser Gly Lys			
215	220	225	
Arg Ile Pro Pro Asp Ile Leu Val Asn Trp Ala Val Gln Ile Ala			
230	235	240	
Arg Gly Met Asn Tyr Leu Leu Asp Glu Ala Ile Val Pro Ile Ile			
245	250	255	
His Arg Asp Leu Lys Ser Ser Asn Ile Leu Ile Leu Gln Lys Val			
260	265	270	
Glu Asn Gly Asp Leu Ser Asn Lys Ile Leu Lys Ile Thr Asp Phe			
275	280	285	
Gly Leu Ala Arg Glu Trp His Arg Thr Thr Lys Met Ser Ala Ala			
290	295	300	
Gly Thr Tyr Ala Trp Met Ala Pro Glu Val Ile Arg Ala Ser Met			
305	310	315	
Phe Ser Lys Gly Ser Asp Val Trp Ser Tyr Gly Val Leu Leu Trp			
320	325	330	
Glu Leu Leu Thr Gly Glu Val Pro Phe Arg Gly Ile Asp Gly Leu			
335	340	345	
Ala Val Ala Tyr Gly Val Ala Met Asn Lys Leu Ala Leu Pro Ile			
350	355	360	
Pro Ser Thr Cys Pro Glu Pro Phe Ala Lys Leu Met Glu Asp Cys			
365	370	375	
Trp Asn Pro Asp Pro His Ser Arg Pro Ser Phe Thr Asn Ile Leu			
380	385	390	
Asp Gln Leu Thr Thr Ile Glu Glu Ser Gly Phe Phe Glu Met Pro			
395	400	405	
Lys Asp Ser Phe His Cys Leu Gln Asp Asn Trp Lys His Glu Ile			
410	415	420	
Gln Glu Met Phe Asp Gln Leu Arg Ala Lys Glu Lys Glu Leu Arg			
425	430	435	
Thr Trp Glu Glu Glu Leu Thr Arg Ala Ala Leu Gln Gln Lys Asn			
440	445	450	
Gln Glu Glu Leu Leu Arg Arg Arg Glu Gln Glu Leu Ala Glu Arg			
455	460	465	
Glu Ile Asp Ile Leu Glu Arg Glu Leu Asn Ile Ile Ile His Gln			
470	475	480	
Leu Cys Gln Glu Lys Pro Arg Val Lys Lys Arg Lys Gly Lys Phe			
485	490	495	
Arg Lys Ser Arg Leu Lys Leu Lys Asp Gly Asn Arg Ile Ser Leu			
500	505	510	
Pro Ser Gly Phe Gln His Lys Phe Thr Val Gln Ala Ser Pro Thr			
515	520	525	
Met Asp Lys Arg Lys Ser Leu Ile Asn Ser Arg Ser Ser Pro Pro			
530	535	540	
Ala Ser Pro Thr Ile Ile Pro Arg Leu Arg Ala Ile Gln Cys Glu			
545	550	555	
Thr Val Ser Lys Thr Trp Gly Arg Ser Ser Val Val Pro Lys Glu			
560	565	570	
Glu Gly Glu Glu Glu Lys Arg Ala Pro Lys Lys Lys Gly Arg			
575	580	585	
Thr Trp Gly Pro Gly Thr Leu Gly Gln Lys Glu Leu Ala Ser Gly			
590	595	600	
Asp Glu Ser Leu Lys Ser Leu Val Asp Gly Tyr Lys Gln Trp Ser			
605	610	615	
Ser Ser Ala Pro Asn Leu Val Lys Gly Pro Arg Ser Ser Pro Ala			
620	625	630	
Leu Pro Gly Phe Thr Ser Leu Met Glu Met Gly Lys Phe Thr Glu			
635	640	645	

Asp Glu Asp Ser Glu Gly Pro Gly Ser Gly Glu Ser Arg Leu Gln  
 650 655 660  
 His Ser Pro Ser Gln Ser Tyr Leu Cys Ile Pro Phe Pro Arg Gly  
 665 670 675  
 Glu Asp Gly Asp Gly Pro Ser Ser Asp Gly Ile His Glu Glu Pro  
 680 685 690  
 Thr Pro Val Asn Ser Ala Thr Ser Thr Pro Gln Leu Thr Pro Thr  
 695 700 705  
 Asn Ser Leu Lys Arg Gly Gly Ala His His Arg Arg Cys Glu Val  
 710 715 720  
 Ala Leu Leu Gly Cys Gly Ala Val Leu Ala Ala Thr Gly Leu Gly  
 725 730 735  
 Phe Asp Leu Leu Glu Ala Gly Lys Cys Gln Leu Leu Pro Leu Glu  
 740 745 750  
 Glu Pro Glu Pro Pro Ala Arg Glu Glu Lys Lys Arg Arg Glu Gly  
 755 760 765  
 Leu Phe Gln Arg Ser Ser Arg Pro Arg Arg Ser Thr Ser Pro Pro  
 770 775 780  
 Ser Arg Lys Leu Phe Lys Lys Glu Glu Pro Met Leu Leu Leu Gly  
 785 790 795  
 Asp Pro Ser Ala Ser Leu Thr Leu Leu Ser Leu Ser Ser Ile Ser  
 800 805 810  
 Glu Cys Asn Ser Thr Arg Ser Leu Leu Arg Ser Asp Ser Asp Glu  
 815 820 825  
 Ile Val Val Tyr Glu Met Pro Val Ser Pro Val Glu Ala Pro Pro  
 830 835 840  
 Leu Ser Pro Cys Thr His Asn Pro Leu Val Asn Val Arg Val Glu  
 845 850 855  
 Arg Phe Lys Arg Asp Pro Asn Gln Ser Leu Thr Pro Thr His Val  
 860 865 870  
 Thr Leu Thr Thr Pro Ser Gln Pro Ser Ser His Arg Arg Thr Pro  
 875 880 885  
 Ser Asp Gly Ala Leu Pro Ser Pro Ser Arg Asp Pro Gly Glu Phe  
 890 895 900  
 Pro Arg Leu Pro Asp Pro Asn Val Val Phe Pro Pro Thr Pro Arg  
 905 910 915  
 Arg Trp Asn Thr Gln Gln Asp Ser Thr Leu Glu Arg Pro Lys Thr  
 920 925 930  
 Leu Glu Phe Leu Pro Arg Pro Arg Pro Ser Ala Asn Arg Gln Arg  
 935 940 945  
 Leu Asp Pro Trp Trp Phe Val Ser Pro Ser His Ala Arg Ser Thr  
 950 955 960  
 Ser Pro Ala Asn Ser Ser Ser Thr Glu Thr Pro Ser Asn Leu Asp  
 965 970 975  
 Ser Cys Phe Ala Ser Ser Ser Ser Thr Val Glu Glu Arg Pro Gly  
 980 985 990  
 Leu Pro Ala Leu Leu Pro Phe Gln Ala Gly Pro Leu Pro Pro Thr  
 995 1000 1005  
 Glu Arg Thr Leu Leu Asp Leu Asp Ala Glu Gly Gln Ser Gln Asp  
 1010 1015 1020  
 Ser Thr Val Pro Leu Cys Arg Ala Glu Leu Asn Thr His Arg Pro  
 1025 1030 1035  
 Ala Pro Tyr Glu Ile Gln Gln Glu Phe Trp Ser  
 1040 1045

<210> 10  
 <211> 357  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474560CD1

<400> 10  
 Met Gln Ile Pro Asp Glu Glu Gly Ile Val Ile Asp Gly Phe Pro  
 1 5 10 15  
 Arg Asp Val Ala Gln Ala Leu Ser Phe Glu Asp Gln Ile Cys Thr  
 20 25 30  
 Pro Asp Leu Val Val Phe Leu Ala Cys Ala Asn Gln Arg Leu Lys  
 35 40 45  
 Glu Arg Leu Leu Lys Arg Ala Glu Gln Gln Gly Arg Pro Asp Asp  
 50 55 60  
 Asn Val Lys Ala Thr Gln Arg Arg Leu Met Asn Phe Lys Gln Asn  
 65 70 75  
 Ala Ala Pro Leu Val Lys Tyr Phe Gln Glu Lys Gly Leu Ile Met  
 80 85 90  
 Thr Phe Asp Ala Asp Arg Asp Glu Asp Glu Val Phe Tyr Asp Ile  
 95 100 105  
 Ser Met Ala Val Asp Asn Lys Leu Phe Pro Asn Lys Glu Ala Ala  
 110 115 120  
 Ala Gly Ser Ser Asp Leu Asp Pro Ser Met Ile Leu Asp Thr Gly  
 125 130 135  
 Glu Ile Ile Asp Thr Gly Ser Asp Tyr Glu Asp Gln Gly Asp Asp  
 140 145 150  
 Gln Leu Asn Val Phe Gly Glu Asp Thr Met Gly Gly Phe Met Glu  
 155 160 165  
 Asp Leu Arg Lys Cys Lys Ile Ile Phe Ile Ile Gly Gly Pro Gly  
 170 175 180  
 Ser Gly Lys Gly Thr Gln Cys Glu Lys Leu Val Glu Lys Tyr Gly  
 185 190 195  
 Phe Thr His Leu Ser Thr Gly Glu Leu Leu Arg Glu Glu Leu Ala  
 200 205 210  
 Ser Glu Ser Glu Arg Ser Lys Leu Ile Arg Asp Ile Met Glu Arg  
 215 220 225  
 Gly Asp Leu Val Pro Ser Gly Ile Val Leu Glu Leu Lys Glu  
 230 235 240  
 Ala Met Val Ala Ser Leu Gly Asp Thr Arg Gly Phe Leu Ile Asp  
 245 250 255  
 Gly Tyr Pro Arg Glu Val Lys Gln Gly Glu Glu Phe Gly Arg Arg  
 260 265 270  
 Ile Gly Asp Pro Gln Leu Val Ile Cys Met Asp Cys Ser Ala Asp  
 275 280 285  
 Thr Met Thr Asn Arg Leu Leu Gln Arg Ser Arg Ser Ser Leu Pro  
 290 295 300  
 Val Asp Asp Thr Thr Lys Thr Ile Ala Lys Arg Leu Glu Ala Tyr  
 305 310 315  
 Tyr Arg Ala Ser Ile Pro Val Ile Ala Tyr Tyr Glu Thr Lys Thr  
 320 325 330  
 Gln Leu His Lys Ile Asn Ala Glu Gly Thr Pro Glu Asp Val Phe  
 335 340 345  
 Leu Gln Leu Cys Thr Ala Ile Asp Ser Ile Ile Phe  
 350 355

&lt;210&gt; 11

&lt;211&gt; 355

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474602CD1

&lt;400&gt; 11

Met Ala Arg Glu Asn Gly Glu Ser Ser Ser Ser Trp Lys Lys Gln  
 1 5 10 15  
 Ala Glu Asp Ile Lys Lys Ile Phe Glu Phe Lys Glu Thr Leu Gly

20	25	30
Thr Gly Ala Phe Ser Glu Val Val Leu Ala Glu Glu Lys Ala Thr		
35	40	45
Gly Lys Leu Phe Ala Val Lys Cys Ile Pro Lys Lys Ala Leu Lys		
50	55	60
Gly Lys Glu Ser Ser Ile Glu Asn Glu Ile Ala Val Leu Arg Lys		
65	70	75
Ile Lys His Glu Asn Ile Val Ala Leu Glu Asp Ile Tyr Glu Ser		
80	85	90
Pro Asn His Leu Tyr Leu Val Met Gln Leu Val Ser Gly Gly Glu		
95	100	105
Leu Phe Asp Arg Ile Val Glu Lys Gly Phe Tyr Thr Glu Lys Asp		
110	115	120
Ala Ser Thr Leu Ile Arg Gln Val Leu Asp Ala Val Tyr Tyr Leu		
125	130	135
His Arg Met Gly Ile Val His Arg Asp Leu Lys Pro Glu Asn Leu		
140	145	150
Leu Tyr Tyr Ser Gln Asp Glu Glu Ser Lys Ile Met Ile Ser Asp		
155	160	165
Phe Gly Leu Ser Lys Met Glu Gly Lys Gly Asp Val Met Ser Thr		
170	175	180
Ala Cys Gly Thr Pro Gly Tyr Val Ala Pro Glu Val Leu Ala Gln		
185	190	195
Lys Pro Tyr Ser Lys Ala Val Asp Cys Trp Ser Ile Gly Val Ile		
200	205	210
Ala Tyr Ile Leu Leu Cys Gly Tyr Pro Pro Phe Tyr Asp Glu Asn		
215	220	225
Asp Ser Lys Leu Phe Glu Gln Ile Leu Lys Ala Glu Tyr Glu Phe		
230	235	240
Asp Ser Pro Tyr Trp Asp Asp Ile Ser Asp Ser Ala Lys Asp Phe		
245	250	255
Ile Arg Asn Leu Met Glu Lys Asp Pro Asn Lys Arg Tyr Thr Cys		
260	265	270
Glu Gln Ala Ala Arg His Pro Trp Ile Ala Gly Asp Thr Ala Leu		
275	280	285
Asn Lys Asn Ile His Glu Ser Val Ser Ala Gln Ile Arg Lys Asn		
290	295	300
Phe Ala Lys Ser Lys Trp Arg Gln Ala Phe Asn Ala Thr Ala Val		
305	310	315
Val Arg His Met Arg Lys Leu His Leu Gly Ser Ser Leu Asp Ser		
320	325	330
Ser Asn Ala Ser Val Ser Ser Ser Leu Ser Leu Ala Ser Gln Lys		
335	340	345
Asp Cys Ala Ser Gly Thr Phe His Ala Leu		
350	355	

&lt;210&gt; 12

&lt;211&gt; 224

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475509CD1

&lt;400&gt; 12

Met Ala Ser Lys Leu Leu Arg Ala Val Ile Leu Gly Pro Pro Gly			
1	5	10	15
Ser Gly Lys Gly Thr Val Cys Gln Arg Ile Ala Gln Asn Phe Gly			
20	25	30	
Leu Gln His Leu Ser Ser Gly His Phe Leu Arg Glu Asn Ile Lys			
35	40	45	
Ala Ser Thr Glu Val Gly Glu Met Ala Lys Gln Tyr Ile Glu Lys			

	50	55	60											
Ser	Leu	Leu	Val	Pro	Asp	His	Val	Ile	Thr	Arg	Leu	Met	Met	Ser
	65	70	75											
Glu	Leu	Glu	Asn	Arg	Arg	Gly	Gln	His	Trp	Leu	Leu	Asp	Gly	Phe
	80	85	90											
Pro	Arg	Thr	Leu	Gly	Gln	Ala	Glu	Ala	Leu	Asp	Lys	Ile	Cys	Glu
	95	100	105											
Val	Asp	Leu	Val	Ile	Ser	Leu	Asn	Ile	Pro	Phe	Glu	Thr	Leu	Lys
	110	115	120											
Asp	Arg	Leu	Ser	Arg	Arg	Trp	Ile	His	Pro	Pro	Ser	Gly	Arg	Val
	125	130	135											
Tyr	Asn	Leu	Asp	Phe	Asn	Pro	Pro	His	Val	His	Gly	Ile	Asp	Asp
	140	145	150											
Val	Thr	Gly	Glu	Pro	Leu	Val	Gln	Gln	Glu	Asp	Asp	Lys	Pro	Glu
	155	160	165											
Ala	Val	Ala	Ala	Arg	Leu	Arg	Gln	Tyr	Lys	Asp	Val	Ala	Lys	Pro
	170	175	180											
Val	Ile	Glu	Leu	Tyr	Lys	Ser	Arg	Gly	Val	Leu	His	Gln	Phe	Phe
	185	190	195											
Arg	Asn	Arg	Arg	Arg	Thr	Lys	Ile	Trp	Pro	Tyr	Val	Tyr	Thr	Thr
	200	205	210											
Phe	Leu	Asn	Lys	Ile	Thr	Pro	Ile	Gln	Ser	Lys	Glu	Ala	Phe	
	215	220												

&lt;210&gt; 13

&lt;211&gt; 502

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475491CD1

&lt;400&gt; 13

Met	Asn	Lys	Met	Lys	Asn	Phe	Lys	Arg	Arg	Phe	Ser	Leu	Ser	Val
1	5	10	15											
Pro	Arg	Thr	Glu	Thr	Ile	Glu	Glu	Ser	Leu	Ala	Glu	Phe	Thr	Glu
	20	25	30											
Gln	Phe	Asn	Gln	Leu	His	Asn	Arg	Arg	Asn	Glu	Asn	Leu	Gln	Leu
	35	40	45											
Gly	Pro	Leu	Gly	Arg	Asp	Pro	Pro	Gln	Glu	Cys	Ser	Thr	Phe	Ser
	50	55	60											
Pro	Thr	Asp	Ser	Gly	Glu	Glu	Pro	Gly	Gln	Leu	Ser	Pro	Gly	Val
	65	70	75											
Gln	Phe	Gln	Arg	Arg	Gln	Asn	Gln	Arg	Arg	Phe	Ser	Met	Glu	Val
	80	85	90											
Arg	Ala	Ser	Gly	Ala	Leu	Pro	Arg	Gln	Val	Ala	Gly	Cys	Thr	His
	95	100	105											
Lys	Gly	Val	His	Arg	Arg	Ala	Ala	Ala	Leu	Gln	Pro	Asp	Phe	Asp
	110	115	120											
Val	Ser	Lys	Arg	Leu	Ser	Leu	Pro	Met	Asp	Ile	Arg	Leu	Pro	Gln
	125	130	135											
Glu	Phe	Leu	Gln	Lys	Leu	Gln	Met	Glu	Ser	Pro	Asp	Leu	Pro	Lys
	140	145	150											
Pro	Leu	Ser	Arg	Met	Ser	Arg	Arg	Ala	Ser	Leu	Ser	Asp	Ile	Gly
	155	160	165											
Phe	Gly	Lys	Leu	Glu	Thr	Tyr	Val	Lys	Leu	Asp	Lys	Leu	Gly	Glu
	170	175	180											
Gly	Thr	Tyr	Ala	Thr	Val	Phe	Lys	Gly	Arg	Ser	Lys	Leu	Thr	Glu
	185	190	195											
Asn	Leu	Val	Ala	Leu	Lys	Glu	Ile	Arg	Leu	Glu	His	Glu	Gly	
	200	205	210											
Ala	Pro	Cys	Thr	Ala	Ile	Arg	Glu	Val	Ser	Leu	Leu	Lys	Asn	Leu

	215	220	225
Lys His Ala Asn	Ile Val Thr Leu His	Asp Leu Ile His Thr	Asp
230	235	240	
Arg Ser Leu Thr	Leu Val Phe Glu Tyr	Leu Asp Ser Asp Leu	Lys
245	250	255	
Gln Tyr Leu Asp	His Cys Gly Asn Leu	Met Ser Met His Asn	Val
260	265	270	
Lys Ile Phe Met	Phe Gln Leu Leu Arg	Gly Leu Ala Tyr Cys	His
275	280	285	
His Arg Lys Ile	Leu His Arg Asp Leu	Lys Pro Gln Asn Leu	Leu
290	295	300	
Ile Asn Glu Arg	Gly Glu Leu Lys Leu	Ala Asp Phe Gly Leu	Ala
305	310	315	
Arg Ala Lys Ser	Val Pro Thr Lys Thr	Tyr Ser Asn Glu Val	Val
320	325	330	
Thr Leu Trp Tyr	Arg Pro Pro Asp Val	Leu Leu Gly Ser Thr	Glu
335	340	345	
Tyr Ser Thr Pro	Ile Asp Met Trp Gly	Val Gly Cys Ile His	Tyr
350	355	360	
Glu Met Ala Thr	Gly Arg Pro Leu Phe	Pro Gly Ser Thr Val	Lys
365	370	375	
Glu Glu Leu His	Leu Ile Phe Arg Leu	Leu Gly Thr Pro Thr	Glu
380	385	390	
Glu Thr Trp Pro	Gly Val Thr Ala Phe	Ser Glu Phe Arg Thr	Tyr
395	400	405	
Ser Phe Pro Cys	Tyr Leu Pro Gln Pro	Leu Ile Asn His Ala	Pro
410	415	420	
Arg Leu Asp Thr	Asp Gly Ile His Leu	Leu Ser Ser Leu Leu	Leu
425	430	435	
Tyr Glu Ser Lys	Ser Arg Met Ser Ala	Glu Ala Ala Leu Ser	His
440	445	450	
Ser Tyr Phe Arg	Ser Leu Gly Glu Arg	Val His Gln Leu Glu	Asp
455	460	465	
Thr Ala Ser Ile	Phe Ser Leu Lys Glu	Ile Gln Leu Gln Lys	Asp
470	475	480	
Pro Gly Tyr Arg	Gly Leu Ala Phe Gln	Gln Pro Gly Arg Gly	Lys
485	490	495	
Asn Arg Arg Gln	Ser Ile Phe		
500			

&lt;210&gt; 14

&lt;211&gt; 791

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2192119CD1

&lt;400&gt; 14

Met Trp Phe Phe Ala	Arg Asp Pro Val	Arg Asp Phe Pro Phe	Glu
1	5	10	15
Leu Ile Pro Glu Pro	Pro Glu Gly Gly	Leu Pro Gly Pro Trp	Ala
20	25	30	
Leu His Arg Gly Arg	Lys Lys Ala Thr	Gly Ser Pro Val	Ser Ile
35	40	45	
Phe Val Tyr Asp Val	Lys Pro Gly Ala	Glu Gln Thr Gln	Val
50	55	60	
Ala Lys Ala Ala Phe	Lys Arg Phe Lys	Thr Leu Arg His	Pro Asn
65	70	75	
Ile Leu Ala Tyr Ile	Asp Gly Leu Glu	Thr Glu Lys Cys	Leu His
80	85	90	
Val Val Thr Glu Ala	Val Thr Pro	Leu Gly Ile Tyr	Leu Lys Ala

95	100	105
Arg Val Glu Ala Gly Gly Leu Lys Glu	Leu Glu Ile Ser Trp	Gly
110	115	120
Leu His Gln Ile Val Lys Ala Leu Ser	Phe Leu Val Asn Asp	Cys
125	130	135
Ser Leu Ile His Asn Asn Val Cys Met	Ala Ala Val Phe Val	Asp
140	145	150
Arg Ala Gly Glu Trp Lys Leu Gly Gly	Leu Asp Tyr Met Tyr	Ser
155	160	165
Ala Gln Gly Asn Gly Gly Pro Pro	Arg Lys Gly Ile Pro	Glu
170	175	180
Leu Glu Gln Tyr Asp Pro Pro Glu Leu	Ala Asp Ser Ser Gly	Arg
185	190	195
Val Val Arg Glu Lys Trp Ser Ala Asp	Met Trp Arg Leu Gly	Cys
200	205	210
Leu Ile Trp Glu Val Phe Asn Gly Pro	Leu Pro Arg Ala Ala	Ala
215	220	225
Leu Arg Asn Pro Gly Lys Ile Pro Lys	Thr Leu Val Pro His	Tyr
230	235	240
Cys Glu Leu Val Gly Ala Asn Pro Lys	Val Arg Pro Asn Pro	Ala
245	250	255
Arg Phe Leu Gln Asn Cys Arg Ala Pro	Gly Gly Phe Met Ser	Asn
260	265	270
Arg Phe Val Glu Thr Asn Leu Phe Leu	Glu Glu Ile Gln Ile	Lys
275	280	285
Glu Pro Ala Glu Lys Gln Lys Phe Phe	Gln Glu Leu Ser Lys	Ser
290	295	300
Leu Asp Ala Phe Pro Glu Asp Phe Cys	Arg His Lys Val Leu	Pro
305	310	315
Gln Leu Leu Thr Ala Phe Glu Phe Gly	Asn Ala Gly Ala Val	Val
320	325	330
Leu Thr Pro Leu Phe Lys Val Gly Lys	Phe Leu Ser Ala Glu	Glu
335	340	345
Tyr Gln Gln Lys Ile Ile Pro Val Val	Val Lys Met Phe Ser	Ser
350	355	360
Thr Asp Arg Ala Met Arg Ile Arg Leu	Leu Gln Gln Met Glu	Gln
365	370	375
Phe Ile Gln Tyr Leu Asp Glu Pro Thr	Val Asn Thr Gln Ile	Phe
380	385	390
Pro His Val Val His Gly Phe Leu Asp	Thr Asn Pro Ala Ile	Arg
395	400	405
Glu Gln Thr Val Lys Ser Met Leu Leu	Leu Ala Pro Lys Leu	Asn
410	415	420
Glu Ala Asn Leu Asn Val Glu Leu Met	Lys His Phe Ala Arg	Leu
425	430	435
Gln Ala Lys Asp Glu Gln Gly Pro Ile	Arg Cys Asn Thr Thr	Val
440	445	450
Cys Leu Gly Lys Ile Gly Ser Tyr Leu	Ser Ala Ser Thr Arg	His
455	460	465
Arg Val Leu Thr Ser Ala Phe Ser Arg	Ala Thr Arg Asp Pro	Phe
470	475	480
Ala Pro Ser Arg Val Ala Gly Val Leu	Gly Phe Ala Ala Thr	His
485	490	495
Asn Leu Tyr Ser Met Asn Asp Cys Ala	Gln Lys Ile Leu Pro	Val
500	505	510
Leu Cys Gly Leu Thr Val Asp Pro Glu	Lys Ser Val Arg Asp	Gln
515	520	525
Ala Phe Lys Ala Ile Arg Ser Phe Leu	Ser Lys Leu Glu Ser	Val
530	535	540
Ser Glu Asp Pro Thr Gln Leu Glu Glu	Val Glu Lys Asp Val	His
545	550	555
Ala Ala Ser Ser Pro Gly Met Gly Gly	Ala Ala Ala Ser Trp	Ala
560	565	570

Gly Trp Ala Val Thr Gly Val Ser Ser Leu Thr Ser Lys Leu Ile  
 575 580 585  
 Arg Ser His Pro Thr Thr Ala Pro Thr Glu Thr Asn Ile Pro Gln  
 590 595 600  
 Arg Pro Thr Pro Glu Gly His Trp Glu Thr Gln Glu Glu Asp Lys  
 605 610 615  
 Asp Thr Ala Glu Asp Ser Ser Thr Ala Asp Arg Trp Asp Asp Glu  
 620 625 630  
 Asp Trp Gly Ser Leu Glu Gln Glu Ala Glu Ser Val Leu Ala Gln  
 635 640 645  
 Gln Asp Asp Trp Ser Thr Gly Gly Gln Val Ser Arg Ala Ser Gln  
 650 655 660  
 Val Ser Asn Ser Asp His Lys Ser Ser Lys Ser Pro Glu Ser Asp  
 665 670 675  
 Trp Ser Ser Trp Glu Ala Glu Gly Ser Trp Glu Gln Gly Trp Gln  
 680 685 690  
 Glu Pro Ser Ser Gln Glu Pro Pro Pro Asp Gly Thr Arg Leu Ala  
 695 700 705  
 Ser Glu Tyr Asn Trp Gly Gly Pro Glu Ser Ser Asp Lys Gly Asp  
 710 715 720  
 Pro Phe Ala Thr Leu Ser Ala Arg Pro Ser Thr Gln Pro Arg Pro  
 725 730 735  
 Asp Ser Trp Gly Glu Asp Asn Trp Glu Gly Leu Glu Thr Asp Ser  
 740 745 750  
 Arg Gln Val Lys Ala Glu Leu Ala Arg Lys Lys Arg Glu Glu Arg  
 755 760 765  
 Arg Arg Glu Met Glu Ala Lys Arg Ala Glu Arg Lys Val Ala Lys  
 770 775 780  
 Gly Pro Met Lys Leu Gly Ala Arg Lys Leu Asp  
 785 790

<210> 15  
 <211> 1651  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474496CD1

<400> 15

Met Ala Gly Gly Arg Gly Ala Pro Gly Arg Gly Arg Asp Glu Pro		
1 5	10	15
Pro Glu Ser Tyr Pro Gln Arg Gln Asp His Glu Leu Gln Ala Leu		
20 25	30	
Glu Ala Ile Tyr Gly Ala Asp Phe Gln Asp Leu Arg Pro Asp Ala		
35 40	45	
Cys Gly Pro Val Lys Glu Pro Pro Glu Ile Asn Leu Val Leu Tyr		
50 55	60	
Pro Gln Gly Leu Thr Gly Glu Glu Val Tyr Val Lys Val Asp Leu		
65 70	75	
Arg Val Lys Cys Pro Pro Thr Tyr Pro Asp Val Val Pro Glu Ile		
80 85	90	
Glu Leu Lys Asn Ala Lys Gly Leu Ser Asn Glu Ser Val Asn Leu		
95 100	105	
Leu Lys Ser Arg Leu Glu Glu Leu Ala Lys Lys His Cys Gly Glu		
110 115	120	
Val Met Ile Phe Glu Leu Ala Tyr His Val Gln Ser Phe Leu Ser		
125 130	135	
Glu His Asn Lys Pro Pro Pro Lys Ser Phe His Glu Glu Met Leu		
140 145	150	
Glu Arg Arg Ala Gln Glu Glu Gln Gln Arg Leu Leu Glu Ala Gln		
155 160	165	

Ala Glu Arg Arg Arg Glu Gln Ala Gln Gln Arg Glu Ile Leu His  
 170 175 180  
 Glu Ile Gln Arg Arg Lys Glu Glu Ile Lys Glu Glu Lys Lys Arg  
 185 190 195  
 Lys Glu Met Ala Lys Gln Glu Arg Leu Glu Ile Ala Ser Leu Ser  
 200 205 210  
 Asn Gln Asp His Thr Ser Lys Lys Asp Pro Gly Gly His Arg Thr  
 215 220 225  
 Ala Ala Ile Leu His Gly Gly Ser Pro Asp Phe Val Gly Asn Gly  
 230 235 240  
 Lys His Arg Ala Asn Ser Ser Gly Arg Ser Arg Arg Glu Arg Gln  
 245 250 255  
 Tyr Ser Val Cys Asn Ser Glu Asp Ser Pro Gly Ser Cys Glu Ile  
 260 265 270  
 Leu Tyr Phe Asn Met Gly Ser Pro Asp Gln Leu Met Val His Lys  
 275 280 285  
 Gly Lys Cys Ile Gly Ser Asp Glu Gln Leu Gly Lys Leu Val Tyr  
 290 295 300  
 Asn Ala Leu Glu Thr Ala Thr Gly Gly Phe Val Leu Leu Tyr Glu  
 305 310 315  
 Trp Val Leu Gln Trp Gln Lys Lys Met Gly Pro Phe Leu Thr Ser  
 320 325 330  
 Gln Glu Lys Glu Lys Ile Asp Lys Cys Lys Lys Gln Ile Gln Gly  
 335 340 345  
 Thr Glu Thr Glu Phe Asn Ser Leu Val Lys Leu Ser His Pro Asn  
 350 355 360  
 Val Val Arg Tyr Leu Ala Met Asn Leu Lys Glu Gln Asp Asp Ser  
 365 370 375  
 Ile Val Val Asp Ile Leu Val Glu His Ile Ser Gly Val Ser Leu  
 380 385 390  
 Ala Ala His Leu Ser His Ser Gly Pro Ile Pro Val His Gln Leu  
 395 400 405  
 Arg Arg Tyr Thr Ala Gln Leu Leu Ser Gly Leu Asp Tyr Leu His  
 410 415 420  
 Ser Asn Ser Val Val His Lys Val Leu Ser Ala Ser Asn Val Leu  
 425 430 435  
 Val Asp Ala Glu Gly Thr Val Lys Ile Thr Asp Tyr Ser Ile Ser  
 440 445 450  
 Lys Arg Leu Ala Asp Ile Cys Lys Glu Asp Val Phe Glu Gln Thr  
 455 460 465  
 Arg Val Arg Phe Ser Asp Asn Ala Leu Pro Tyr Lys Thr Gly Lys  
 470 475 480  
 Lys Gly Asp Val Trp Arg Leu Gly Leu Leu Leu Ser Leu Ser  
 485 490 495  
 Gln Gly Gln Glu Cys Gly Glu Tyr Pro Val Thr Ile Pro Ser Asp  
 500 505 510  
 Leu Pro Ala Asp Phe Gln Asp Phe Leu Lys Cys Val Cys Leu Asp  
 515 520 525  
 Asp Lys Glu Arg Trp Ser Pro Gln Gln Leu Leu Lys His Ser Phe  
 530 535 540  
 Ile Asn Pro Gln Pro Lys Met Pro Leu Val Glu Gln Ser Pro Glu  
 545 550 555  
 Asp Ser Glu Gly Gln Asp Tyr Val Glu Thr Val Ile Pro Ser Asn  
 560 565 570  
 Arg Leu Pro Ser Ala Ala Phe Phe Ser Glu Thr Gln Arg Gln Phe  
 575 580 585  
 Ser Arg Tyr Phe Ile Glu Phe Glu Glu Leu Gln Leu Leu Gly Lys  
 590 595 600  
 Gly Ala Phe Gly Ala Val Ile Lys Val Gln Asn Lys Leu Asp Gly  
 605 610 615  
 Cys Cys Tyr Ala Val Lys Arg Ile Pro Ile Asn Pro Ala Ser Arg  
 620 625 630  
 Gln Phe Arg Arg Ile Lys Gly Glu Val Thr Leu Leu Ser Arg Leu

	635	640	645
His His Glu Asn Ile Val Arg Tyr Tyr		Asn Ala Trp Ile Glu Arg	
650	655	660	
His Glu Arg Pro Ala Gly Pro Gly Thr		Pro Pro Pro Asp Ser Gly	
665	670	675	
Pro Leu Ala Lys Asp Asp Arg Ala Ala		Arg Gly Gln Pro Ala Ser	
680	685	690	
Asp Thr Asp Gly Leu Asp Ser Val Glu		Ala Ala Ala Pro Pro Pro	
695	700	705	
Ile Leu Ser Ser Ser Val Glu Trp Ser		Thr Ser Gly Glu Arg Ser	
710	715	720	
Ala Ser Ala Arg Phe Pro Ala Thr Gly		Pro Gly Ser Ser Asp Asp	
725	730	735	
Glu Asp Asp Asp Glu Asp Glu His Gly		Gly Val Phe Ser Gln Ser	
740	745	750	
Phe Leu Pro Ala Ser Asp Ser Glu Ser		Asp Ile Ile Phe Asp Asn	
755	760	765	
Glu Asp Glu Asn Ser Lys Ser Gln Asn		Gln Asp Glu Asp Cys Asn	
770	775	780	
Glu Lys Asn Gly Cys His Glu Ser Glu		Pro Ser Val Thr Thr Glu	
785	790	795	
Ala Val His Tyr Leu Tyr Ile Gln Met		Glu Tyr Cys Glu Lys Ser	
800	805	810	
Thr Leu Arg Asp Thr Ile Asp Gln Gly		Leu Tyr Arg Asp Thr Val	
815	820	825	
Arg Leu Trp Arg Leu Phe Arg Glu Ile		Leu Asp Gly Leu Ala Tyr	
830	835	840	
Ile His Glu Lys Gly Met Ile His Arg		Asp Leu Lys Pro Val Asn	
845	850	855	
Ile Phe Leu Asp Ser Asp Asp His Val		Lys Ile Gly Asp Phe Gly	
860	865	870	
Leu Ala Thr Asp His Leu Ala Phe Ser		Ala Asp Ser Lys Gln Asp	
875	880	885	
Asp Gln Thr Gly Asp Leu Ile Lys Ser		Asp Pro Ser Gly His Leu	
890	895	900	
Thr Gly Met Val Gly Thr Ala Leu Tyr		Val Ser Pro Glu Val Gln	
905	910	915	
Gly Ser Thr Lys Ser Ala Tyr Asn Gln		Lys Val Asp Leu Phe Ser	
920	925	930	
Leu Gly Ile Ile Phe Phe Glu Met Ser		Tyr His Pro Met Val Thr	
935	940	945	
Ala Ser Glu Arg Ile Phe Val Leu Asn		Gln Leu Arg Asp Pro Thr	
950	955	960	
Ser Pro Lys Phe Pro Glu Asp Phe Asp		Asp Gly Glu His Ala Lys	
965	970	975	
Gln Lys Ser Val Ile Ser Trp Leu Leu		Asn His Asp Pro Ala Lys	
980	985	990	
Arg Pro Thr Ala Thr Glu Leu Leu Lys		Ser Glu Leu Leu Pro Pro	
995	1000	1005	
Pro Gln Met Glu Glu Ser Glu Leu His		Glu Val Leu His His Thr	
1010	1015	1020	
Leu Thr Asn Val Asp Gly Lys Ala Tyr		Arg Thr Met Met Ala Gln	
1025	1030	1035	
Ile Phe Ser Gln Arg Ile Ser Pro Ala		Ile Asp Tyr Thr Tyr Asp	
1040	1045	1050	
Ser Asp Ile Leu Lys Gly Asn Phe Ser		Ile Arg Thr Ala Lys Met	
1055	1060	1065	
Gln Gln His Val Cys Glu Thr Ile Ile		Arg Ile Phe Lys Arg His	
1070	1075	1080	
Gly Ala Val Gln Leu Cys Thr Pro Leu		Leu Leu Pro Arg Asn Arg	
1085	1090	1095	
Gln Ile Tyr Glu His Asn Glu Ala Ala		Leu Phe Met Asp His Ser	
1100	1105	1110	

Gly Met Leu Val Met Leu Pro Phe Asp Leu Arg Ile Pro Phe Ala  
 1115 1120 1125  
 Arg Tyr Val Ala Arg Asn Asn Ile Leu Asn Leu Lys Arg Tyr Cys  
 1130 1135 1140  
 Ile Glu Arg Val Phe Arg Pro Arg Lys Leu Asp Arg Phe His Pro  
 1145 1150 1155  
 Lys Glu Leu Leu Glu Cys Ala Phe Asp Ile Val Thr Ser Thr Thr  
 1160 1165 1170  
 Asn Ser Phe Leu Pro Thr Ala Glu Ile Ile Tyr Thr Ile Tyr Glu  
 1175 1180 1185  
 Ile Ile Gln Glu Phe Pro Ala Leu Gln Glu Arg Asn Tyr Ser Ile  
 1190 1195 1200  
 Tyr Leu Asn His Thr Met Leu Leu Lys Ala Ile Leu Leu His Cys  
 1205 1210 1215  
 Gly Ile Pro Glu Asp Lys Leu Ser Gln Val Tyr Ile Ile Leu Tyr  
 1220 1225 1230  
 Asp Ala Val Thr Glu Lys Leu Thr Arg Arg Glu Val Glu Ala Lys  
 1235 1240 1245  
 Phe Cys Asn Leu Ser Leu Ser Ser Asn Ser Leu Cys Arg Leu Tyr  
 1250 1255 1260  
 Lys Phe Ile Glu Gln Lys Gly Asp Leu Gln Asp Leu Met Pro Thr  
 1265 1270 1275  
 Ile Asn Ser Leu Ile Lys Gln Lys Thr Gly Ile Ala Gln Leu Val  
 1280 1285 1290  
 Lys Tyr Gly Leu Lys Asp Leu Glu Glu Val Val Gly Leu Leu Lys  
 1295 1300 1305  
 Lys Leu Gly Ile Lys Leu Gln Val Leu Ile Asn Leu Gly Leu Val  
 1310 1315 1320  
 Tyr Lys Val Gln Gln His Asn Gly Ile Ile Phe Gln Phe Val Ala  
 1325 1330 1335  
 Phe Ile Lys Arg Arg Gln Arg Ala Val Pro Glu Ile Leu Ala Ala  
 1340 1345 1350  
 Gly Gly Arg Tyr Asp Leu Leu Ile Pro Gln Phe Arg Gly Pro Gln  
 1355 1360 1365  
 Ala Leu Gly Pro Val Pro Thr Ala Ile Gly Val Ser Ile Ala Ile  
 1370 1375 1380  
 Asp Lys Ile Ser Ala Ala Val Leu Asn Met Glu Glu Ser Val Thr  
 1385 1390 1395  
 Ile Ser Ser Cys Asp Leu Leu Val Val Ser Val Gly Gln Met Ser  
 1400 1405 1410  
 Met Ser Arg Ala Ile Asn Leu Thr Gln Lys Leu Trp Thr Ala Gly  
 1415 1420 1425  
 Ile Thr Ala Glu Ile Met Tyr Asp Trp Ser Gln Ser Gln Glu Glu  
 1430 1435 1440  
 Leu Gln Glu Tyr Cys Arg His His Glu Ile Thr Tyr Val Ala Leu  
 1445 1450 1455  
 Val Ser Asp Lys Glu Gly Ser His Val Lys Val Lys Ser Phe Glu  
 1460 1465 1470  
 Lys Glu Arg Gln Thr Glu Lys Arg Val Leu Glu Thr Glu Leu Val  
 1475 1480 1485  
 Asp His Val Leu Gln Lys Leu Arg Thr Lys Val Thr Asp Glu Arg  
 1490 1495 1500  
 Asn Gly Arg Glu Ala Ser Asp Asn Leu Ala Val Gln Asn Leu Lys  
 1505 1510 1515  
 Gly Ser Phe Ser Asn Ala Ser Gly Leu Phe Glu Ile His Gly Ala  
 1520 1525 1530  
 Thr Val Val Pro Ile Val Ser Val Leu Ala Pro Glu Lys Leu Ser  
 1535 1540 1545  
 Ala Ser Thr Arg Arg Tyr Glu Thr Gln Val Gln Thr Arg Leu  
 1550 1555 1560  
 Gln Thr Ser Leu Ala Asn Leu His Gln Lys Ser Ser Glu Ile Glu  
 1565 1570 1575  
 Ile Leu Ala Val Asp Leu Pro Lys Glu Thr Ile Leu Gln Phe Leu

	1580	1585	1590
Ser Leu Glu Trp Asp Ala Asp Glu Gln Ala Phe Asn Thr Thr Val			
1595	1600	1605	
Lys Gln Leu Leu Ser Arg Leu Pro Lys Gln Arg Tyr Leu Lys Leu			
1610	1615	1620	
Val Cys Asp Glu Ile Tyr Asn Ile Lys Val Glu Lys Lys Val Ser			
1625	1630	1635	
Val Leu Phe Leu Tyr Ser Tyr Arg Asp Asp Tyr Tyr Arg Ile Leu			
1640	1645	1650	
Phe			

<210> 16  
 <211> 752  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1834248CD1

<400> 16

Met Ser Ser Arg Thr Val Leu Ala Pro Gly Asn Asp Arg Asn Ser			
1	5	10	15
Asp Thr His Gly Thr Leu Gly Ser Gly Arg Ser Ser Asp Lys Gly			
20	25	30	
Pro Ser Trp Ser Ser Arg Ser Leu Gly Ala Arg Cys Arg Asn Ser			
35	40	45	
Ile Ala Ser Cys Pro Glu Glu Gln Pro His Val Gly Asn Tyr Arg			
50	55	60	
Leu Leu Arg Thr Ile Gly Lys Gly Asn Phe Ala Lys Val Lys Leu			
65	70	75	
Ala Arg His Ile Leu Thr Gly Arg Glu Val Ala Ile Lys Ile Ile			
80	85	90	
Asp Lys Thr Gln Leu Asn Pro Ser Ser Leu Gln Lys Leu Phe Arg			
95	100	105	
Glu Val Arg Ile Met Lys Gly Leu Asn His Pro Asn Ile Val Lys			
110	115	120	
Leu Phe Glu Val Ile Glu Thr Glu Lys Thr Leu Tyr Leu Val Met			
125	130	135	
Glu Tyr Ala Ser Ala Gly Glu Val Phe Asp Tyr Leu Val Ser His			
140	145	150	
Gly Arg Met Lys Glu Lys Glu Ala Arg Ala Lys Phe Arg Gln Ile			
155	160	165	
Val Ser Ala Val His Tyr Cys His Gln Lys Asn Ile Val His Arg			
170	175	180	
Asp Leu Lys Ala Glu Asn Leu Leu Leu Asp Ala Glu Ala Asn Ile			
185	190	195	
Lys Ile Ala Asp Phe Gly Phe Ser Asn Glu Phe Thr Leu Gly Ser			
200	205	210	
Lys Leu Asp Thr Phe Cys Gly Ser Pro Pro Tyr Ala Ala Pro Glu			
215	220	225	
Leu Phe Gln Gly Lys Lys Tyr Asp Gly Pro Glu Val Asp Ile Trp			
230	235	240	
Ser Leu Gly Val Ile Leu Tyr Thr Leu Val Ser Gly Ser Leu Pro			
245	250	255	
Phe Asp Gly His Asn Leu Lys Glu Leu Arg Glu Arg Val Leu Arg			
260	265	270	
Gly Lys Tyr Arg Val Pro Phe Tyr Met Ser Thr Asp Cys Glu Ser			
275	280	285	
Ile Leu Arg Arg Phe Leu Val Leu Asn Pro Ala Lys Arg Cys Thr			
290	295	300	
Leu Glu Gln Ile Met Lys Asp Lys Trp Ile Asn Ile Gly Tyr Glu			

Gly	Glu	Glu	Leu	305	Lys	Pro	Tyr	Thr	Glu	Pro	Glu	Glu	Asp	Phe	Gly	310	
				320											325	330	
Asp	Thr	Lys	Arg	Ile	Glu	Val	Met	Val	Gly	Met	Gly	Tyr	Thr	Arg	335	340	345
Glu	Glu	Ile	Lys	Glu	Ser	Leu	Thr	Ser	Gln	Lys	Tyr	Asn	Glu	Val	350	355	360
Thr	Ala	Thr	Tyr	Leu	Leu	Leu	Gly	Arg	Lys	Thr	Glu	Glu	Gly	Gly	365	370	375
Asp	Arg	Gly	Ala	Pro	Gly	Leu	Ala	Leu	Ala	Arg	Val	Arg	Ala	Pro	380	385	390
Ser	Asp	Thr	Thr	Asn	Gly	Thr	Ser	Ser	Ser	Lys	Gly	Thr	Ser	His	395	400	405
Ser	Lys	Gly	Gln	Arg	Ser	Ser	Ser	Ser	Thr	Tyr	His	Arg	Gln	Arg	410	415	420
Arg	His	Ser	Asp	Phe	Cys	Gly	Pro	Ser	Pro	Ala	Pro	Leu	His	Pro	425	430	435
Lys	Arg	Ser	Pro	Thr	Ser	Thr	Gly	Glu	Ala	Glu	Leu	Lys	Glu	Glu	440	445	450
Arg	Leu	Pro	Gly	Arg	Lys	Ala	Ser	Cys	Ser	Thr	Ala	Gly	Ser	Gly	455	460	465
Ser	Arg	Gly	Leu	Pro	Pro	Ser	Ser	Pro	Met	Val	Ser	Ser	Ala	His	470	475	480
Asn	Pro	Asn	Lys	Ala	Glu	Ile	Pro	Glu	Arg	Arg	Lys	Asp	Ser	Thr	485	490	495
Ser	Thr	Pro	Asn	Asn	Leu	Pro	Pro	Ser	Met	Met	Thr	Arg	Arg	Asn	500	505	510
Thr	Tyr	Val	Cys	Thr	Glu	Arg	Pro	Gly	Ala	Glu	Arg	Pro	Ser	Leu	515	520	525
Leu	Pro	Asn	Gly	Lys	Glu	Asn	Ser	Ser	Gly	Thr	Pro	Arg	Val	Pro	530	535	540
Pro	Ala	Ser	Pro	Ser	Ser	His	Ser	Leu	Ala	Pro	Pro	Ser	Gly	Glu	545	550	555
Arg	Ser	Arg	Leu	Ala	Arg	Gly	Ser	Thr	Ile	Arg	Ser	Thr	Phe	His	560	565	570
Gly	Gly	Gln	Val	Arg	Asp	Arg	Arg	Ala	Gly	Gly	Gly	Gly	Gly	Gly	575	580	585
Gly	Val	Gln	Asn	Gly	Pro	Pro	Ala	Ser	Pro	Thr	Leu	Ala	His	Glu	590	595	600
Ala	Ala	Pro	Leu	Pro	Ala	Gly	Arg	Pro	Arg	Pro	Thr	Thr	Asn	Leu	605	610	615
Phe	Thr	Lys	Leu	Thr	Ser	Lys	Leu	Thr	Arg	Arg	Val	Ala	Asp	Glu	620	625	630
Pro	Glu	Arg	Ile	Gly	Gly	Pro	Glu	Val	Thr	Ser	Cys	His	Leu	Pro	635	640	645
Trp	Asp	Gln	Thr	Glu	Thr	Ala	Pro	Arg	Leu	Leu	Arg	Phe	Pro	Trp	650	655	660
Ser	Val	Lys	Leu	Thr	Ser	Ser	Arg	Pro	Pro	Glu	Ala	Leu	Met	Ala	665	670	675
Ala	Leu	Arg	Gln	Ala	Thr	Ala	Ala	Ala	Arg	Cys	Arg	Cys	Arg	Gln	680	685	690
Pro	Gln	Pro	Phe	Leu	Leu	Ala	Cys	Leu	His	Gly	Gly	Ala	Gly	Gly	695	700	705
Pro	Glu	Pro	Leu	Ser	His	Phe	Glu	Val	Glu	Val	Cys	Gln	Leu	Pro	710	715	720
Arg	Pro	Gly	Leu	Arg	Gly	Val	Leu	Phe	Arg	Arg	Val	Ala	Gly	Thr	725	730	735
Ala	Leu	Ala	Phe	Arg	Thr	Leu	Val	Thr	Arg	Ile	Ser	Asn	Asp	Leu	740	745	750
Glu	Leu																

&lt;211&gt; 501

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71584520CD1

&lt;400&gt; 17

Met	Pro	Phe	Gly	Cys	Val	Thr	Leu	Gly	Asp	Lys	Lys	Asn	Tyr	Asn
1					5				10			15		
Gln	Pro	Ser	Glu	Va	l	Thr	Asp	Arg	Tyr	Asp	Leu	Gly	Gln	Val
					20					25			30	
Lys	Thr	Glu	Glu	Phe	Cys	Glu	Ile	Phe	Arg	Ala	Lys	Asp	Lys	Thr
					35				40			45		
Thr	Gly	Lys	Leu	His	Thr	Cys	Lys	Lys	Phe	Gln	Lys	Arg	Asp	Gly
					50					55		60		
Arg	Lys	Val	Arg	Lys	Ala	Ala	Lys	Asn	Glu	Ile	Gly	Ile	Leu	Lys
					65				70			75		
Met	Val	Lys	His	Pro	Asn	Ile	Leu	Gln	Leu	Val	Asp	Val	Phe	Val
					80				85			90		
Thr	Arg	Lys	Glu	Tyr	Phe	Ile	Phe	Leu	Glu	Leu	Ala	Thr	Gly	Arg
					95				100			105		
Glu	Val	Phe	Asp	Trp	Ile	Leu	Asp	Gln	Gly	Tyr	Tyr	Ser	Glu	Arg
					110				115			120		
Asp	Thr	Ser	Asn	Va	l	Val	Arg	Gln	Val	Leu	Glu	Ala	Val	Ala
					125					130		135		
Leu	His	Ser	Leu	Lys	Ile	Val	His	Arg	Asn	Leu	Lys	Leu	Glu	Asn
					140				145			150		
Leu	Val	Tyr	Tyr	Asn	Arg	Leu	Lys	Asn	Ser	Lys	Ile	Val	Ile	Ser
					155				160			165		
Asp	Phe	His	Leu	Ala	Lys	Leu	Glu	Asn	Gly	Leu	Ile	Lys	Glu	Pro
					170				175			180		
Cys	Gly	Thr	Pro	Glu	Tyr	Leu	Ala	Pro	Glu	Val	Val	Gly	Arg	Gln
					185				190			195		
Arg	Tyr	Gly	Arg	Pro	Val	Asp	Cys	Trp	Ala	Ile	Gly	Val	Ile	Met
					200				205			210		
Tyr	Ile	Leu	Leu	Ser	Gly	Asn	Pro	Pro	Phe	Tyr	Glu	Glu	Val	Glu
					215				220			225		
Glu	Asp	Asp	Tyr	Glu	Asn	His	Asp	Lys	Asn	Leu	Phe	Arg	Lys	Ile
					230				235			240		
Leu	Ala	Gly	Asp	Tyr	Glu	Phe	Asp	Ser	Pro	Tyr	Trp	Asp	Asp	Ile
					245				250			255		
Ser	Gln	Ala	Ala	Lys	Asp	Leu	Val	Thr	Arg	Leu	Met	Glu	Val	Glu
					260				265			270		
Gln	Asp	Gln	Arg	Ile	Thr	Ala	Glu	Glu	Ala	Ile	Ser	His	Glu	Trp
					275				280			285		
Ile	Ser	Gly	Asn	Ala	Ala	Ser	Asp	Lys	Asn	Ile	Lys	Asp	Gly	Val
					290				295			300		
Cys	Ala	Gln	Ile	Glu	Lys	Asn	Phe	Ala	Arg	Ala	Lys	Trp	Lys	
					305				310			315		
Ala	Val	Arg	Val	Thr	Thr	Leu	Met	Lys	Arg	Leu	Arg	Ala	Pro	Glu
					320				325			330		
Gln	Ser	Ser	Thr	Ala	Ala	Ala	Gln	Ser	Ala	Ser	Ala	Thr	Asp	Thr
					335				340			345		
Ala	Thr	Pro	Gly	Ala	Ala	Gly	Gly	Ala	Thr	Ala	Ala	Ala	Ser	
					350				355			360		
Gly	Ala	Thr	Ser	Ala	Pro	Glu	Gly	Asp	Ala	Ala	Arg	Ala	Ala	Lys
					365				370			375		
Ser	Asp	Asn	Val	Ala	Pro	Ala	Asp	Arg	Ser	Ala	Thr	Pro	Ala	Thr
					380				385			390		
Asp	Gly	Ser	Ala	Thr	Pro	Ala	Thr	Asp	Gly	Ser	Val	Thr	Pro	Ala
					395				400			405		

Thr Asp Gly Ser Ile Thr Pro Ala Thr Asp Gly Ser Val Thr Pro  
 410 415 420  
 Ala Thr Asp Arg Ser Ala Thr Pro Ala Thr Asp Gly Arg Ala Thr  
 425 430 435  
 Pro Ala Thr Glu Glu Ser Thr Val Pro Thr Thr Gln Ser Ser Ala  
 440 445 450  
 Met Leu Ala Thr Lys Ala Ala Ala Thr Pro Glu Pro Ala Met Ala  
 455 460 465  
 Gln Pro Asp Ser Thr Ala Pro Glu Gly Ala Thr Gly Gln Ala Pro  
 470 475 480  
 Pro Ser Ser Lys Gly Glu Glu Ala Ala Gly Tyr Ala Gln Glu Ser  
 485 490 495  
 Gln Arg Glu Glu Ala Ser  
 500

<210> 18  
 <211> 346  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7475538CD1

<400> 18  
 Met Asp Gln Tyr Cys Ile Leu Gly Arg Ile Gly Glu Gly Ala His  
 1 5 10 15  
 Gly Ile Val Phe Lys Ala Lys His Val Glu Thr Gly Glu Ile Val  
 20 25 30  
 Ala Leu Lys Lys Val Ala Leu Arg Arg Leu Glu Asp Gly Phe Pro  
 35 40 45  
 Asn Gln Ala Leu Arg Glu Ile Lys Ala Leu Gln Glu Met Glu Asp  
 50 55 60  
 Asn Gln Tyr Val Val Gln Leu Lys Ala Val Phe Pro His Gly Gly  
 65 70 75  
 Gly Phe Val Leu Ala Phe Glu Phe Met Leu Ser Asp Leu Ala Glu  
 80 85 90  
 Val Val Arg His Ala Gln Arg Pro Leu Ala Gln Ala Gln Val Lys  
 95 100 105  
 Ser Tyr Leu Gln Met Leu Leu Lys Gly Val Ala Phe Cys His Ala  
 110 115 120  
 Asn Asn Ile Val His Arg Asp Leu Lys Pro Ala Asn Leu Leu Ile  
 125 130 135  
 Ser Ala Ser Gly Gln Leu Lys Ile Ala Asp Phe Gly Leu Ala Arg  
 140 145 150  
 Val Phe Ser Pro Asp Gly Ser Arg Leu Tyr Thr His Gln Val Ala  
 155 160 165  
 Thr Arg Trp Tyr Arg Ala Pro Glu Leu Leu Tyr Gly Ala Arg Gln  
 170 175 180  
 Tyr Asp Gln Gly Val Asp Leu Trp Ser Val Gly Cys Ile Met Gly  
 185 190 195  
 Glu Leu Leu Asn Gly Ser Pro Leu Phe Pro Gly Lys Asn Asp Ile  
 200 205 210  
 Glu Gln Leu Cys Tyr Val Leu Arg Ile Leu Gly Thr Pro Asn Pro  
 215 220 225  
 Gln Val Trp Pro Glu Leu Thr Glu Leu Pro Asp Tyr Asn Lys Ile  
 230 235 240  
 Ser Phe Lys Glu Gln Val Pro Met Pro Leu Glu Glu Val Leu Pro  
 245 250 255  
 Asp Val Ser Pro Gln Ala Leu Asp Leu Leu Gly Gln Phe Leu Leu  
 260 265 270  
 Tyr Pro Pro His Gln Arg Ile Ala Ala Ser Lys Ala Leu Leu His  
 275 280 285

Gln Tyr Phe Phe Thr Ala Pro Leu Pro Ala His Pro Ser Glu Leu  
 290 295 300  
 Pro Ile Pro Gln Arg Leu Gly Gly Pro Ala Pro Lys Ala His Pro  
 305 310 315  
 Gly Pro Pro His Ile His Asp Phe His Val Asp Arg Pro Leu Glu  
 320 325 330  
 Glu Ser Leu Leu Asn Pro Glu Leu Ile Arg Pro Phe Ile Leu Glu  
 335 340 345  
 Gly

<210> 19  
 <211> 4224  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2890544CB1

<400> 19  
 cggcggttag ccccgacgg cagcaggagg ccgagggcggg agcgcgcggg gctgaggcgg 60  
 cggcgccggc ggcggaaagg ggctgggtgg tggagcggga gggaggctga ggaggctccc 120  
 cctgcgggac gggcgcgggg acggctccgg gggcggggg ccggggcccg ggccggctcg 180  
 cggcggtttt atgatgaccc ggctgcgggg ccccagatct cgtctccctcc gccgcctcc 240  
 cacggcagcc ccagctcgcg gctgagaacc agacacacccg gcggtatggc atcacagctg 300  
 caagtgtttt cgccccatc agtgcgtcg agtgccttc gcaatgcgaa gaaaactgaaa 360  
 atagagccct ctggctggga tgttcagga cagatgatca acgacaaaata ttataccac 420  
 agcaaaaccc tcccagccac acaagggcaa gccaactcct ctcaccaggat agcaaatttc 480  
 aacatccctg cttacgacca gggcctccctc ctcccagctc ctgcagtggc gcatattgtt 540  
 gtaacagccg ctgatagtc gggcaagtgc gctacatcaa cttccaaag cagccagacc 600  
 ctgactcaca gaagcaacgt ttcttcgtt gagccatataaaaaatgtgg attgaaacga 660  
 aaaagtgagg aagtgcacag caacggtagt gtgcagatca tagagaaca tcccccttc 720  
 atgctgcaaa acaggactt ggtgggtgc gtcgcaccaa ccaccactgt gaccacaaag 780  
 attagcgtt ccacggaga aggggattac cagctggtc acatgatgc ctttgcgtct 840  
 atgaccaata gctatgaagt cttggaggatc ctggccggg ggacattttgg acagggtggct 900  
 aagtgcgttga agaggagcac caaggaaattt gtggcttattaaatctt gaaaccacccc 960  
 tcctatgcca gacaaggaca gatttgcgtt agcatcctt cccgcctaag cagtggaaat 1020  
 gctgtgagt ataatttgcgtt ccgttcatac gagtgcgtt agcataagaa tcacacctgc 1080  
 ctgttttttgcgtt aaatgttggc gcagaactta tatgattttc taaagcaaaa caaaatttagc 1140  
 ccactgcac tcaagtacat cagaccaatc ttgcgcggg tggccacagc ctgtatgaag 1200  
 ctcaagagtc ttggctgtat ccacgcgtac cttaaaggctt aaaaatcatcat gtcgggtgtat 1260  
 ccagttcgcgcc agccctaccg agtgcggatc attgactttt gttctgcgtac tcacgtttcc 1320  
 aaagctgtgt gctcaaccta cttacgtca cttactaca gagctcctga aattattctt 1380  
 gggttaccat ttgtgaacg tattgatatc ttgtcactgg gtcgtgtatc agtcgagctg 1440  
 ttctggat ggccttta tcctgggtgtatc tcaaatatc atcagatccg ttatatttca 1500  
 caaacacaag gcttgcgcgc tgaatatctt ctcgtgcgc gaaacaaaac aaccaggat 1560  
 ttcaacagag atcctaattt ggggtacccca ctgtggaggc ttaagacacc tgaagaacat 1620  
 gaactggaga ctggaaataaa atcaaaaaggaa gtcggaaatc acattttaa ttgcgttagat 1680  
 gacatggctc aggtgaatat gtctacagac ctggaggggaa cagacatgtt ggcagagaaag 1740  
 gcaagccgaa gagaatacat tgatctgtt aagaaaatgc tcacaatttgc tgcagataag 1800  
 agaattaccc ctctaaaaac tcttaaccat cagttgttca caatgactca cttttggat 1860  
 tttcacata gcaatcatgt taagtcttgcgtt tccttgcgttca tggagatctg caagcggagg 1920  
 gttcacatgt atgatcactgt gaggctcgtt aagacttccct tcactacaca tggccccc 1980  
 aatacaagca caaatctacat catgagcttc agcaatcgc tcaatcactgt gcaaatcag 2040  
 gcaatgttc tagcttcacat ttctactgtca gcaatgttca ctctttcttgcgttca 2100  
 gatgtctcactt tactaaactt ccacgttgcgtt ttttgcgttca tggatccat catctgcgtc accatgttctt 2160  
 ggagttggcc agcagggtgt ttcccttgcgtt cttggaaacca cccagatgg cactcagaca 2220  
 gatccattcc aacagacatt tatagtatgt ccacctgcgtt ttcaacttgcgtt actacaagca 2280  
 acaacaaaggc attctggatt ccctgtgagg atggataatgc tgcgttccat tggatccat 2340  
 gcaccagctg ctcagccact acagatttcgc tcaaggatgtt tcacgcgggg aagctgtaca 2400  
 ccactaatgg tagcaactt ccacccctcaaa gtagccacca tcacaccgcgca gtatgcgttgcgtt 2460  
 ccctttactc tgagctgcgc agccggccgg ccggcgctgg ttgaacagac tgccgtgttca 2520

ctcgaggcgt	ggcctggagg	gactcagcaa	attctcctgc	cttcaacttg	gcaacagttg	2580
cctgggttag	ctctacacaa	ctctgtccag	cccacagcaa	tgattccaga	ggccatgggg	2640
agtggacagc	agctagctga	ctggaggaat	gcccactctc	atggcaacca	gtacagcact	2700
atcatgcagc	agccatcctt	gctgactaac	catgtgacat	tggccactgc	tcagcccttg	2760
aatgttggtg	ttgcccattgt	tgtcagacaa	caacaatcca	gttccctccc	ttcgaagaag	2820
aataagcagt	cagotccagt	ctcttccaag	tcctctctag	atgttctgc	ttcccaagtc	2880
tattctctgg	ttgggagcag	tcccctccgc	accacatctt	tttataattc	cttggtccct	2940
gtccaagatc	agcatcagcc	catcatcatt	ccagatactc	ccagccctcc	tgtgagttgc	3000
atcaactatcc	gaagtgcac	tgatgagggaa	gaggacaaca	aatacaagcc	cagtagctct	3060
ggactgaagc	caaggctctaa	tgtcatcagt	tatgtcactg	tcaatgattc	tccagactct	3120
gacttttctt	tgagcagccc	tatattccat	gataccctga	gtgctctccg	aggcaataatg	3180
ggatccgttt	ttggggggcc	tggcagagg	gtggcagatg	gcaactggcc	ccgcactatc	3240
atttgtcctc	cactgaaaac	tcagcttgg	gactgcactg	tagcaacccca	ggccctcagg	3300
ctccctgagca	ataagactaa	gccagtcgt	tcagtgatgt	ggcagtcattc	tggatgttgt	3360
atcaccccca	cagggtatcg	agctcaacgc	ggggggacca	gtgcagcaca	accactcaat	3420
cttagccaga	accagcagtc	atcggcggt	ccaaacctcac	agggagagaag	cagcaacccca	3480
gccccccgca	ggcagcaggc	gtttgtggcc	cctctctccc	aagcccccta	caccttccag	3540
catggcagcc	cgctacactc	gacagggcac	ccacaccccttg	ccccggccccc	tgctcacctg	3600
ccaagccagg	ctcatctgt	tacgtatgt	gccccgactt	ctgctgctgc	actgggctca	3660
accagctcca	ttgctcatct	tttctccca	cagggttctt	caaggcatgc	tgcagcctat	3720
accactcacc	ctagcacttt	ggtgcaccag	gtccctgtca	gtgttggcc	cagccttcctc	3780
acttctgcca	gctgtggccccc	tgctcagtag	caacaccagt	ttgcccaccca	atcctacatt	3840
gggtcttccc	gaggctcaac	aatttacact	ggatacccg	tgagtccctac	caagatcagc	3900
cagtatccct	acttataatgtt	ggtgagcatg	agggaggagg	aatcatggct	accttctct	3960
ggccctgct	tcttaatatt	gggtatgg	gagatccttc	tttaccctct	tgaaatttct	4020
tagccagcaa	cttggtctgc	agggggccac	tgaagcagaa	ggtttttttc	tggggaaacc	4080
tgtctcagt	ttgactgcat	ttttgtatgc	ttctcaaagt	tggccctatt	tttaaatcca	4140
tatttttgt	acgtcattcg	gtactggaga	gtcaatgcca	tttctgcgtc	cagcagaatg	4200
gctcgatcc	ctgaatttgc	cctg				4224

<210> 20

<211> 1736

<212> DNA

<213> *Homo sapiens*

<220>

<221> misc\_feature

<223> Incyte ID No: 7472693CB1

<400> 20

cgccggcgccgg	cgagagcgaa	agaggaaact	gcagaggagg	aagctgcgcc	gcagcccgag	60
ccgccccggca	tccccggccgc	ctctgcgccc	gcccgcgcc	cccgccgccc	cctcccccagc	120
gegccccccgg	ccgctctcc	ggccgcgcgt	cgtcgccat	ggcccgggag	aacggcgaga	180
gcagctctc	ctggaaaaag	caagctgaag	acatcaagaa	gatttcgag	ttcaaagaga	240
ccctcggaac	cggggcctt	tccgaagtgg	tttagctga	agagaaggca	actggcaagc	300
tcttgctgt	gaagtgtatc	cctaagaagg	cgctgaaggg	caaggaaagc	agcatagaga	360
atgagatagc	cgtcctgaga	agatgttgc	atgaaaat	tgttgcctg	gaagacattt	420
atgaaaaggcc	aaatcacctg	tacttggtca	tgcagctgtt	gtccgggtga	gagctgtttg	480
acggatagat	ggagaaaaagg	ttttatacag	agaaggatgc	cagcactctg	atccggccaa	540
tcttggacgc	cgttactat	ctccacagaa	tgggcacatgt	ccacagagac	ctcaaggccg	600
aaaatcttct	gtactacagt	caagatgagg	agtccaaaat	aatgtatcgt	gactttggat	660
tgtcaaaaaat	ggagggcaaa	ggagatgtga	tgtccactgc	ctgtggaaact	ccaggctatg	720
tcgctcttga	agtctctcgcc	cagaaacctt	acagcaaagc	cgttgcgtc	tggccatcg	780
gagtgattgc	ctacatcttgc	ctctgcggct	accctcttt	ttatgtgaa	aatgactcca	840
agctctttga	gcagatcctc	aaggcggaaat	atgagtttga	ctctccctac	tgggatgaca	900
tctccgactc	tgcaaaaagac	ttcattcgg	acctgtatgg	gaaggacccg	aataaaagat	960
acacgtgtga	gcaggcagct	cggcacccat	ggatcgctgg	tgacacagcc	ctcaacaaaa	1020
acatccacga	gtccgtcage	gcccagatcc	ggaaaaactt	tgccaagagc	aatggagac	1080
aaggattttaa	tgccacggcc	gtctgtggagac	atatggaaaa	actacacctc	ggcagcagcc	1140
tggacagttc	aaaatgcagaat	gtttcggagca	gcctcgat	ggccagccaa	aaagactgt	1200
cgttatgtagc	aaaaccagaa	tccctcagct	gacactgt	acgagcctgg	ggtggagagg	1260
aggggagccgg	catctggcga	gcacccctgt	tttgcacaggc	gtttctata	cttaatccca	1320
tgtcatgcga	cccttaggact	tttttaaca	tgtatctact	ggccgggggtg	cagtggctca	1380

cgccgtgtaat cccaaacactt tgggaggcgt aggccaggagg actgtttgag ttcaggagtt 1440  
 ttaagaccag cctgaccaac atggtaaac cccatctca ctaaaatata aaaatttagcc 1500  
 gggtgtgtgt gcgagcacct gtaatgtcag ctacttggga ggctgaggca ggagaatcac 1560  
 ttgaacccag gaagcggagg ttgcaatgag ctgagatcac accactgcac tccagcctgg 1620  
 gtacagacatt gagactccct ctcaaaaaaa aaaggaaat cattaaacac tcgttggaaacc 1680  
 ctaggtattt catattccat ttacggttt ggaatccagg gctcaagtcc tcgcag 1736

<210> 21  
 <211> 1824  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3107952CB1

<400> 21  
 cacggcttcc ggtgtcatgg ctgcttgaag tcccggaggt cggtgaggcg gctgcaggc 60  
 cctccctgcg gagccgcgtgg tccggctggc ggagatgtga ccgcgggccc ggccggcctg 120  
 cctcaggcgt cgcgtcagct cccgtgtccg tgcccttaac ccacaccgt ggcgggatcc 180  
 ggctgcgcct ggggcgcgga gcccgcgtt tttctggagg cttcgggccc gctgtggcag 240  
 gtacagagcc gtctgggtag cggctcctcc gcctcgggtt atcgggttcg ctgctgcggc 300  
 aaccctggc cggcccccgg cggccctcaag cagttcttc cgcgggac caccgggct 360  
 gcccgcctg cggccgagta tgggttccgc aaagagaggg cggcgtcggc acagttgcag 420  
 ggtcacagaa acatgtgac tttgtatggc gtgtttacaa tccactttc tccaaatgtg 480  
 ccatcacgct gtctgttgc tgaactcctg gatgtcagtg tttcggaaat gctcttat 540  
 tccagtcacc agggttttc catgtggatg atacagcatt gtgcccggaa tgggttggag 600  
 gcccttgc ttcttcatca tgaggcgtat gtccatgcgg acctcaaaacc acgttaacata 660  
 ttgtggagtg cagagaatga atgtttaaa ctcattgact ttggacttag ctccaaagaa 720  
 ggaatcagg atgtaaagta tattcagaca gacgggtatc gggctccaga agcagaattg 780  
 caaaaattgtct tggcccaggc tggcctgcag agtatacag aatgtaccc agctgttgat 840  
 ctgtggagcc taggaatcat tttactggaa atgttctcag gaatggaaact gaaacatata 900  
 gtcagatctc aggaatggaa ggcaaacagt tctgttattt ttgatcacat atttgcagg 960  
 aaagcagtgg tgaatgcgc aattccagcc tatcacctaa gagacccat caaaaagcatg 1020  
 cttcatgtat atccaaaggc aagaattctt gctgaaatgg cattgtcag cccattctt 1080  
 agcattccctt ttgcccctt tatttgcattt ctggcatgc ttcccactcc agtgcataa 1140  
 ctgctgaatg tgctggatga tgattttttt gagaatggaa aggaatatga agatgttgta 1200  
 gaagatgtaa aagaggagtg tcaaaaatata gacgggtgg tatctctact tttccaaag 1260  
 ggaaatcctg gcagaggaca agtctttgtt gatgtcggaa atgctggta ttccaaagct 1320  
 ggcgcagaaat tactgactgg aaggatgttt gatggaaatgtt ttgttggc tacattctac 1380  
 ccgctgatgtt cttttttcc tttttttttt cttgggttat tccacatatg aatgcaggac 1440  
 tccccctta ccatttttaag aaggatctttt atacattttt ttaatcttac taatgtcag 1500  
 ccattgcaca agcagtgtact gctgtcata catttggcac tgatgttgc aagacctctc 1560  
 agtatacat tgagggttt taggcattcc atgtgggca ccctttttt tgccggagag 1620  
 cagggttgc tcttcgtat gtagccaaa aaaaatcttaa ttatccatg gatcatgaag 1680  
 caaggatgaa taatatcatg tcttggtaaa tactaacaaa ttgtttaggt ttgtgacat 1740  
 catttacagg ttatccctt atgtt 1800  
 1824

<210> 22  
 <211> 2201  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5544420CB1

<400> 22  
 atgaaccgat acacaaccat gagacagttt ggggacggca cgtatgggag tttgtttat 60  
 ggcaagagta atgaatccgg ggagctgggtt gccatcaaaa ggatgttgc gaaatgttat 120  
 tcttggatg aatgcattttt cttggatggaa gtaatgttgc tgaagaaact taatcatgcc 180  
 aatgttattttaa aattgttgc gaaatgttgc aatgttattt tatattttttt 240

tatataaaaaaaaatcata tcaattaatg aaagacagga acaagggtt ccctgaatca 300  
 gtcataatcata tcaaatattg caagggctgg ctttatcca taaacatgg 360  
 tttttcata gggacatgaa accagaaaac ttgcgttgc tgggtccaga gttgtgaaa 420  
 attgtctgatt ttggacttgc aagagaatta aggtcacagc caccatacac tgattatgt 480  
 tctaccagat ggtatcgat ccctgaagtt ttactgagat cttcaatccat tagttctcc 540  
 attgtatgtt gggctgtgg aagatcatg gctgaactct atatgttaag gccactttc 600  
 ccaggacaa gtgaggtcga tgaatctt aaaatttgc aagtttagg gactccaa 660  
 aaaagtact ggccagaagg ataccagct gcatcctcta tgaacttccg tttcccaag 720  
 tgtgttcata taaacttaaa aactcttatt cccaatgcga gtaatgaagc tattcagtc 780  
 atgaccgaaa tggtgaattt ggtatccaaag aaacgaccga cagcaagcca ggcattgaaa 840  
 caccatattt tcgaatgtt tcaggtatta ggccttgcg caaatcatct ggaatcaaaa 900  
 cagtcttaa ataaggact gcaaccatc gaatcaaaagc catcttagt ttaggttagag 960  
 cctaaggctc tgccggatatt aatcgatcag gttgtggac aaccggcc aaaaacttagc 1020  
 cagcagccac tgcaatccat tcagccgcca cagaacctga gctgtccagca acctccaaag 1080  
 caacagagtc aggagaaacc gccacaaacg ctatcccg gcatcgtaa aacatgcca 1140  
 actaaggccaa atggcacact gagtcataaa agtggtaga ggcgttgggg tcagactatc 1200  
 ttcaagtctg gagatagctg ggaagagttt gaggactatg atttccggagc ctccattcc 1260  
 aagaaggccaa gcatgggtt tttaaagaa aaaaggaaaa aagattctcc atttccggctt 1320  
 ccagagccag taccctcagg ctccaaaccac tcgacagggg aaaacaagag ctacactgct 1380  
 gttaacttccc taaaatctga ttccgaattt tcaactgctc caacctctaa acagactac 1440  
 ttgaaacaat caagatatct tccagggtgtt aatcccaaga aggtgcctt gatagccagt 1500  
 gggaaaggaaaa taaacccca cacttggagc aaccagttat tcccaagtc actgggaccc 1560  
 gtggggcag aacttgcattt caaaaggagc aatgcagaag aaaaacttgg aagttatgtt 1620  
 acttacaatc agtcaggata tatttcattt tttctcaaaa aagaagtgc gtcagctggc 1680  
 cagaggatcc acttcgttcc tctcaatgcg acggcttcg aatataccgtt gacacacaaaa 1740  
 actggcggg ggcagtttc aggacgtact tataatccata cagcaaaaaa cctaaatatt 1800  
 gtgaaccgtc cacagccat tccctcgtt cacgggagga cagactgggtt ggccttgc 1860  
 ggaggccacc ggttaggatc tatgggtgtt aaccctacag cattgtccg tagagtacgt 1920  
 gcaagttcct tgaccctggg aaatgtctac aaatgtctat ttctactgag ttctggaaaga 1980  
 aatatgcaaa agtgggtact tggaaaggccaa aacatcatcc cctatttttac ttatccaa 2040  
 gaaatgcatt tcttagcat catttccac agtgggtata tatgggttagg atgttacaaa 2100  
 gtattgaata aactatttgc caaagtatga agtatttgcata tacaatttata ataaatagta 2160  
 aatccaataa gaacccttaa aaaaaacaa cttccagaaaa a 2201

&lt;210&gt; 23

&lt;211&gt; 2974

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472832CB1

&lt;400&gt; 23

gcggatccc gcggatcttag aaccagggtt ccccccggggc ccccccggcg ggtcccggt 60  
 gggctccagg cggccggtcc cggccctccc cccatggca cggcccccctt ttatcccgcc 120  
 gggctccctg getctccccc gcccgggtct cttccggccc cggccggcct agagctgcag 180  
 tcggcccccac cgctactgccc ccagatcccg gcccgggtt cgggggtctt ctccatc 240  
 cagatggggc tgaccctggca gttcggtctg ttggccggcc cttccggactt ggtcatgt 300  
 aagcagctgg cctgttccat cgtggaccag aagttccctg agtgggtttt ctacggcctt 360  
 tacgacaaga ttctgtttt caaacatgac cccacgtcg ccaacccctt gcaagctgg 420  
 cgctcgccg gagacatcca ggaggccgac ctgggtggagg tgggtgtgtc ggcctccggc 480  
 accttcgagg acttccagat ccggccgcac gcccacccgg tgcactctta tggggccct 540  
 gccttcgtt atcaactgccc ggagatgtctc ttccggcttgc tgcggccaggc cctcaagtgc 600  
 gatggctcg ggtgtactt ccacaaggcgc ttgcgttca gcatcccaaa caactgtatg 660  
 gggcccccga aacggccgct gtcatccacg ttctggccca gtggccactc ggtggccctc 720  
 ggcacccctcg agtccctgccc ctgcacccggc gaagaggtgtc gccgtacccac caccggactc 780  
 ctggccctggc gtcccccgtt atcccttcc ttctcttcttgc cttccatcgta tacggggccgc 840  
 cccatttgagc tggacaagat gctgtctcc aagggtcaagg tgccgcacac ctccctcatc 900  
 cacagctata cacggccac cgtttggccag gtttgcagaaga aactccctaa gggcccttcc 960  
 cggcaggggcc tggcaatgcgaa agactgcag tttactgtc acaaagctg cggccacccgc 1020  
 gtccttaatg actgcctggg ggaggccctt atcaatggag atgtgcgtt gggaggcc 1080  
 accgatttca gcgaggctga caagagcgcc cttcatggatc agtcagagga ctccgggttc 1140

atccctggct cccactcaga gaatgcgctc cacGCCAGTG aggaggagga aggcgaggga 1200  
 ggcaaggccc agagctccct ggggtacatc cccctaataa ggggtgtgcg atcggtgcga 1260  
 cacacgacgc gaaatccag caccacgctg cgggagggtt ggggtgtca ttacagcaac 1320  
 aaggacacgc tgagaaagcg gcactattgg cgcctggact gcaagtgtat cacgcttcc 1380  
 cagaacaaca cgaaccaacag atactataag gaaattccgc tgcagaaat ctcacggtg 1440  
 gagtccgccc agaacttcag ccttgcgcg cgggcacca acccacactg ctttgagatc 1500  
 gtcactgcca atgccaccta cttcgtggc gagatgcctg gcgggactcc ggggtggcc 1560  
 agtgggcagg gggctgaggc cgcccgcccc tggagacag ccatccgcca gcccctgatc 1620  
 cccgtcatcc ttcaggacgc acccagcgc ccaggccacg cgccccacag acaagcttc 1680  
 ctgagcatcc ctgtgtccaa cagtcagatc caagagaatg tggacattgc cactgtctac 1740  
 cagatcttc ctgacgaatg gctgggtca gggcagttt gagtggctta tggagaaaaa 1800  
 caccggaga aaggccggg cgtggcagtt aaggtcatcg acaaactgcg cttccctacc 1860  
 aagcaggaga gcccgcctg gaatgaatg gccattctgc agagcctgcg gcatcccg 1920  
 atcgtgaacc tggagtgcattt gttcggagac cctggagaaatg tggggagaag 1980  
 ctgcatgggg acatgttggg gatgatcctg ctcaggatgaa agggccggct gcctgagcgc 2040  
 ctcaccaatg tcctcatcac ccagatcctg gtggcttgcg acacatttca cttcaagaac 2100  
 attgtccact gtgacttgaa accagaaaaac gtgttgccttgcg catcagcaga cccatttc 2160  
 caggtaagc tgggtgactt tggcttgcg cgcattcatcg gcgagaaatg gttccgccc 2220  
 tcagtgggtgg gcaacccggc ctacctggca cccgagggtgc tgctcaacca gggctacaac 2280  
 cgctcgctgg acatgttggc agtggggcgtg atcatgtacg tcagcctcag cggcaccc 2340  
 ccttcacacg aggatggaga catcaatgac cagatccaga acgcccctt catgtaccc 2400  
 gccagcccat ggagccacat ctcaactggc gccatttgcg tcattcaacaa cctgctgc 2460  
 gtaagatgc gcaaaacgta cagctggac aaatcttca gccacccctg gtacaggag 2520  
 taccagacgt ggctggaccc cccgagatc gaggggaaaga tgggagagcgt atacatcacg 2580  
 catgagatg acgacgcgcg ctggagacgat tttgcagcag agcatccgt gcctgggtct 2640  
 gggctgccc cggacaggga tctcgggtgg gcctgtccac caagagcca cgacatgc 2700  
 gggctggcgg agcgcattcag tgggttgcg ggtctgtgc cctcgtccag ctgctgcct 2760  
 ccacagcgg tcttcacagg atcccagca tgaactgttc tagggaaatg ggttcc 2820  
 ccaaactgga tgggacacgt ggggagttggg gtggggggag ctatttccaa gggccctccc 2880  
 tgggttccca gcaattaaaaa cggactcatc tctggcccat ggccttgatc tcaaaaaaaa 2940  
 aaaaaaaaaaaa aaaaaaaaaacgg cggtcgcaag cttt 2974

&lt;210&gt; 24

&lt;211&gt; 3648

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1551456CB1

&lt;400&gt; 24

gggtacgacg cggatcaacta ttacggcgca gtgtgttggaa aagtgttca agagatctta 60  
 agctggaggc accaggtctg aattccagac tccctccac caccacact tcacccatca 120  
 ctggagcatg accacagacc cattcaggga ggctggcgga ctcttcatcc tggacagtcc 180  
 cttactgtat gtcaaaagctg agaatgaagc ggagagcattc agacagagga gttggggaaa 240  
 cgtcgccag gccaaggctt cttagaagtgg gatattctgg aaataatgca aagagagctg 300  
 gaccattcat ctttggtccc cgtctggca actcaccggt gccaaggata gtgcagtgtt 360  
 tggcgaggaa agatggcact gatgacttct atcagctgaa gatcctgacc ctggaggaga 420  
 ggggggacca aggcatagag agccagggaaag agccggcagg caagatgtc ctgcacaccc 480  
 agtactcaat gctgtctctc ctgcacacgc aggtatggcgt ggtgcaccac caccggctct 540  
 tccaggaccc cacctgtgaa atcgttgagg acacacatc cagccggatg gtttggggaaa 600  
 tgaagaagcg catctgcctc gtctggact gcctctgtgc tcatgacttc agcgataaga 660  
 ccgtctgaccc catcaacccctt cagcactactg tcatcaaggaa gaagagatc agcgagagg 720  
 agactgttgtt aatcttctac gacgtggtcc gcgtgggtgg ggcctgtcactt cagaaaaata 780  
 tcgtgcacag agacctgaag ctggggaaaca tgggtgttcaaa caagagggata catcggtataa 840  
 ccatcaccaat cttctgcctc gggaaacatc tgggtgttcaaa gggggacactg ctgaaggacc 900  
 agagagggag ccctgcctac atcgttcccg acgtgttccat cggccggccg taccgtggca 960  
 agcccaatgtt catgtggggcc ctggggcgtgg tgccttccat catgtgtat ggcctgttcc 1020  
 cttcttacgtt cagcatcccg caggagctt tccgttccat caaggctgtcc gaggatccat 1080  
 ttctgtggaa tggacgggtt tctgtggaaaca cctgtgttccat cttccggaaatg ctgtgggtcc 1140  
 ttggggggggccat gcaaccccttgc gggccggccg acgttccatca ggcctcttccat ggcatttccat 1200  
 catcatggca gttccctgttca tctctgttccat ggccttttccat gacattgtat 1260

accaaatgag	caatgcggat	agctcccaagg	aggcgaagg	gacggaggag	tgctccca	1320
acgagtttga	gaactacatg	cgtcagcagc	tgctgtggc	cgaggagaag	agctccatcc	1380
atgacgcccc	gagctgggt	cccaagcggc	agttcggcag	cgcacccac	gtgcacggc	1440
tgggccacga	cgcacagccc	atgacctct	tggacacggc	catcctggc	cagegttacc	1500
tgcgaaata	acagcctcg	ccggggccac	cagcactgct	gccacttct	ccagccccag	1560
ccaaaggcgt	ggctgtcagg	gctgggcct	gtagtgctgg	actctcccg	gccacaata	1620
ggacaggcga	gggacaggga	cagcccaggt	cacacgtgg	gtcagcagag	gtaccacgaa	1680
gctacccctt	gggtatgtt	ctcgatttt	tggttttaa	atctgagaag	cctagataac	1740
taatctgtt	ttaatcacg	tgttttaat	tacccctgtc	tcttttaacca	tgctgtctct	1800
ggactgagca	agagggagga	ggggacctgc	tcacccact	ccagggcct	ccccagcggc	1860
caccaactga	cctggggcgc	tgctccccc	agtccaaata	agtcgaag	gcagctcgct	1920
gcagggccca	gagcggagct	cccccctcc	ctgcttccc	agggccctgc	cacagcctc	1980
ttccgtccct	ctcttctga	tccaggcccc	tcagtc	tttggaaaa	ccttcaccc	2040
atcttaaacc	gaactcaa	atatttattt	ttttaccata	coaacttctc	tcccatctct	2100
aggtggctca	gtccatggcc	actccctg	cccagcctgg	ctggacagca	aggaatccac	2160
agcccacacg	tgagctccct	cctcacc	aggcagggaa	gcccctctg	ccagtccctg	2220
tccccttca	gcccacc	ccctctctgc	tgccgtt	gggaggcctt	tctagacctg	2280
getctttctc	tcccg	gtggcttctc	tgaggt	tacacgcgc	ttaacctgtt	2340
cccttctcta	tccttcccc	tggtactg	ctcacgtg	cttccagtg	gaagggccc	2400
atgggttgg	ctg	ggccgtgagc	ggggctg	tgacgttcc	cctagctac	2460
tcttgtt	aggggatgtt	ggaaacat	cccagt	ttgcctcata	atagatgtgg	2520
tgactctcc	ggtagaccc	agcaagg	ctccatgg	gtgagggact	caggagaatt	2580
gtagggattt	ggggacc	ctgc	ttgagaac	ccctgt	cttttgagcc	2640
gagatttga	agtggatg	ctgttgc	aaaatgt	tccacagg	atgccccctc	2700
cccttgcct	tactggactt	ggccctg	gatgc	aaagacc	ccccagaggc	2760
ctacccccc	tatgtc	gagagg	gtgtccc	caggc	tggccctg	2820
ggcccttct	gc	gtccc	ggggagg	ctgtt	cagagtgt	2880
gcccgttcc	tcccctaaa	aagctg	actgt	ac	at	2940
cctcctt	gtt	ccat	aaaaatgg	ccac	ccagc	3000
caggggt	ca	tct	ccatcc	cact	actcat	3060
ctgcctc	ca	gg	ccc	tttgc	ctgagc	3120
agcctgc	tc	gg	cc	act	cttgc	3180
gggatgc	ctt	ttt	tttgc	tttgc	gtgggg	3240
tccggcct	gg	cc	tttgc	tttgc	ggaaat	3300
ggccgtcg	gc	cc	tttgc	tttgc	ctggg	3360
gcgttgtt	tct	tt	tttgc	tttgc	tgct	3420
cagactgg	c	tca	tttgc	tttgc	ggac	3480
agaggaggat	gggtc	act	tttgc	tttgc	tatt	3540
aaacggaaaa	agtc	act	tttgc	tttgc	catggg	3600
cttc	act	tttgc	tttgc	tttgc	tttgc	3648

<210> 25  
<211> 4719  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2589355CB1

```

<400> 25
gcggagccgc cgtcgccccaa gccaagccgc gtcgccaacc ctcccgccccg cccgcgcgtcc 60
tgtccgcgt gtctagcagc ggggcccagc atggtcatgg cgatggccc gaggcacttg 120
cagcgcgggc cggtccgggt ggggttctac gacatcgagg gcacgctggg caagtggcaa 180
cttcgcgtgt gtgaagctgg ggcggccaccg gatcaccaag acggagggtgg caataaaaat 240
aatcgataag ttcagccgt ggatgcatgt gaactttagaaaatctacc gagaagtaca 300
aataatgaaa atgttagacc accctcacat aatcaaactt ttcaggtaa tggagaccaa 360
aagtatgttg taccttgtga cagaatatgc caaaaatggaa gaaatttttg atactcttc 420
taatcatggc cggttaaatg agtctgaagc caggcgaaaa ttctggcaaa tcctgtctgc 480
tgttgattat tgtcatggtc ggaagattgt gcaccgtgac ctcaaaagctg aaaatctct 540
gctggataac aacatgaata tcaaaaatagc agatttcggt ttggaaatt tctttaaaag 600
tggtaactg ctggcaacat ggtgtggcag ccccccttat gcagccccag aagtcttga 660
agggcagcag tatgaaggac cacagctgga catctggagt atgggagtttgc ttctttatgt 720

```

ccttgcgtgt ggagcttcgc ccttgcgtgg accgactt ccaattttga ggcagagggt 780  
 tctggaaaggga agattccggaa ttccgttattt catgtcagaa gattgcgagc accttatccg 840  
 aaggatgttg gtccttagacc catccaaacg gctaaccata gcccaaatacaggagcataa 900  
 atggatgctc atagaagttc ctgtccagag acctgttctc tatccacaag agcaagaaaa 960  
 tgagccatcc atcggagagt ttaatgagca ggttctgca ctgtatgcaca gccttggaaat 1020  
 agatcagcag aaaaccattt agtctttgca gaacaagagc tataaccact ttgctgccat 1080  
 ttatcttgc ttgggtggc gcctgaaatc acatcgaggc agttttccag tggagcagag 1140  
 acttgcgtgc cgcctggc ggcctagcac cattgctgag caaacagttt ccaaggcaca 1200  
 gactgtgggg ctcccagtga ccatgcattt accgaacatg aggctgtgc gatctgcct 1260  
 cctcccccag gcatccaaacg tggaggcctt ttcatatttca gcacatcggt gtcaggcgg 1320  
 agctgcattt atggaagaag agtgtgtggc cactccaaag gtcaatggcgt gtcctgcttga 1380  
 ccctgtgcctt cctgtcctgg tgccgaaggg atgcccagtca ctgcccagca acatgtatgg 1440  
 gaccccttattt gacgaaggggc tggagacaga aggagaggcc gaggaaagacc ccgctcatgc 1500  
 ctttgaggca tttcgttca caccgcggc gcagagacgg cacactctgt cagaagtgc 1560  
 caatcaactt gtcgtgttgc ctggggcagg gaaaattttc tccatgtatgc acagccccctc 1620  
 ccttgacagt gtggactctg agtatgtat ggggtctgtt cagagggacc tgaactttct 1680  
 ggaagacaac ctttccctta aggacatcat tttagccat cgccttcac cccgcatgc 1740  
 atctcccttc ataaggcttgc gacttacca cccagccatg caggctctga gtcctccagaa 1800  
 acgagaggcc cacaacagggt ctccaggatg cttcagagag ggccgcagag catcagatac 1860  
 ctccctcacc cagggattt tagattttt acaacatctt cagaatctgg cttagaacc 1920  
 aggaatttcta gagttgaaca aagtgcagtt ttgttatgaa caaataggac cgaggaggc 1980  
 ccctaaccctg ggcggcggc ctccctcagct ccaggacctt gctaggatgc gccctcagga 2040  
 agaaggtttct cagcagcagg aaagcttc cacttccctt gccagcgtgc atccccagct 2100  
 gtccccacggc cagggcttgc agacccagta cctgcagcac agactccaga agcccgac 2160  
 tctgtcaaaag gcccagaaca cctgtcagct ttatttgcattt gaaaccacccg ggagccttga 2220  
 gcagcagctg cagggatc cggccggcggc ctccctcagct ccaggacctt gctaggatgc gccctcagga 2280  
 actgcaggcc tattttatc agatgcagatc agcagagacgc tccatccccac agccaagtca 2340  
 gcaagctgccc ctccccccggc aggagacttcc accgccttctt cagcaggccc caccgttcc 2400  
 cctgacccttgc cccctgagcc ccgttccgttgc gccttcccttcc gaggatgc aatacagccc 2460  
 tttccctcagc cagttccaaag agatgcagct tcagccccctt ccctccactt ccggcccccc 2520  
 ggctgtctt cctctgccttcc cgcagctaca gcagcagcagc ccggccacccg caccaccccc 2580  
 tcaccacca cgcacggccggc gagctgcccccc agccccctt cagttccctt atcagacttg 2640  
 tgagctgccc agcgtcttgc cccctgccc agactatccc actccctgtc agtattctgt 2700  
 ggtatggagcc cagcagacgc acctaaccggg gccagactgtt cccagaagcc caggactgca 2760  
 agaggcccccc tccagcttgc acccaacttgc cctctctgttgc ttttgcatttgc tctttgatttgc 2820  
 tggaaatgttgc tggctgttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 2880  
 gcacaggaaat tgggtgggtt caggttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 2940  
 ctttttaattt taaagtttat ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3000  
 aacttggaaatc agagggtctt gctgggggtt atgttgcatttgc ttttgcatttgc ttttgcatttgc 3060  
 caaaggttttc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3120  
 cgagtggagcc aagctcttgc gggcagacttgc gacaaatgttgc ttttgcatttgc ttttgcatttgc 3180  
 cccagggtctt atgcaggattt acatccgttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3240  
 aagggtgtgt gctatttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3300  
 agcaaccaca tatttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3360  
 aataaaaggaa agtctgttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3420  
 tcccccaaggcc ccatcacaat ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3480  
 cagctgttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3540  
 ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3600  
 acgttggaaatc gacaggagac cccggccaccgttgc ggaggcagggttgc ggcaagaaac tcaagaacgc 3660  
 atcaagagca ccaggcccttgc gccaggaaatc acaggcttgc ttttgcatttgc ttttgcatttgc 3720  
 ctgttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3780  
 ctatcttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3840  
 ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3900  
 agtggtagtgc cgggttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 3960  
 caacagtttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4020  
 atgtggaaatc cggatgttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4080  
 attacagcaatc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4140  
 ccaaacatcaatc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4200  
 taaaatttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4260  
 ttggacagtttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4320  
 gcccattttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4380  
 taaaggcataatc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4440  
 tcctttttgttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc ttttgcatttgc 4500

ggccatgcga gcccagctcc taccaacgta ggttaacttga gcagtccctg ttgctggcca 4560  
 gagactgcct ggtcgccagc gctcaacatg ggtgccagga tgcttcgcag aggcaactgtg 4620  
 ctcacgggtt gacttgggtg cagtgggaaa gggcagtgtg gggactgtca tttttgtat 4680  
 ttataacaca cgtaaaaatc aggaagaatg aataagctt 4719  
  
 <210> 26  
 <211> 1651  
 <212> DNA  
 <213> Homo sapiens  
  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 4357117CB1  
  
 <400> 26  
 atgaggattt gttgttttagt gaaaaaccaa cagccccctgg gagccaccat caagcgccac 60  
 gagatgacag gggacatctt ggtggccagg atcatccacg gtgggctggc ggagagaagt 120  
 ggtttctat atgctggaga caaactggta gaagtgaatg gagtttcagt tgaggactg 180  
 gaccctgaac aagtgtatcca tattctggcc atgtctcgag gcacaatcat gtcaaggtg 240  
 gttccagtct ctgaccctcc tggtaatagc cagcagatgt tatacgtccg tgccatgact 300  
 gagaactggc cccaggagga tcccgacatc ccctgcattt acgctggatt gcctttccag 360  
 aagggggaca tccctccagat tggaccatg aatgtatgccc tctggggca gggccggaaaa 420  
 atctcagacc ctgttacactt cgctgggctt gtcccttcta accaccttctt gaagagggaa 480  
 caacggaaat tctggggc tcagccgtac cagcctcaca cctgcctcaa gtcaacctca 540  
 gacaaggagg agtttgggtt ctacgggtc aagtctttt taggttaggtt cagcccgctg 600  
 catggcagtg tggctgtcac cggcagctgc tacagtgcag tgggtcccc ttacggagg 660  
 gtgggtgggtt accagcagc accttcagac aagtaccggc tcatagtgtt cataggaccc 720  
 tctgggtttt gaggtaatgta gctcagaaga caacttattt aatttatcc cagccatttt 780  
 caaaagtgtcg tggccacac tactcgact aaaaagaggat acgaaatgaa tggcgctgag 840  
 tatcaactatg tggccaaatg aacattttt aacctcatat atagtcacag gatgctggag 900  
 tatgggtgggtt acaaaaggcca cctgtatggc actagtgtttt atgctgttca aacagtccctt 960  
 gtcgaaggaa agatctgtgtt catggaccta gaccccttggg atatccaagg gtttgcacc 1020  
 catgaactga agccctatgtt catattata aagccatcg atatgaggt tatgaaacaa 1080  
 tctcgaaaaa atgccaaggat tattactgac tactatgtgg acatgaagtt caaggatgaa 1140  
 gacccatcaag agatggaaaa tttagcccaa agaatggaaa ctcagtttgg ccaatttttt 1200  
 gatcatgtga ttgtgaatgta cagcttgcac gatgtatgtt cccagttttt gtctggccata 1260  
 cagaaggctc aggaggagcc tcagttggta ccagcaacat ggatttccctc agataactgag 1320  
 tctcaatgag ctccatgtttaatgtggta gttttacacat tggatcccttgc atacagcgat 1380  
 ccatagttgc aatctaaaac aacagtattt gacccatattt aatgtgtaca actttaaaag 1440  
 tgcagcaatt tattatatttta tcttatttttta aaaaaattttt tattgtatgg ttatgtgggtt 1500  
 acctattttta acttaattttt ttttccctta cctcatatgc agctgtggta gaaatatgaa 1560  
 taatgttaag tcaactgagta tgagaacctt tcgcagattt cacatgtatct ttttaagatt 1620  
 taaaataaaga gtttccctaa ataaaaaaaaa a 1651  
  
 <210> 27  
 <211> 3141  
 <212> DNA  
 <213> Homo sapiens  
  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5511992CB1  
  
 <400> 27  
 atggagccctt ccagagcgct tctcggtgtc ctacgcggcg cggccgtgc cgccccggcg 60  
 gggggggatg gggcaggggc cggggcccgag gaggaggagg aggaggaggaa ggaggccggc 120  
 gcggcggtgg gccccggggc gctgggtgtc gacgcgcggc tgccctactt gacggccgtg 180  
 ttccgtatcg aggccggggc cgaggacgag ctgaccctgc ggctggccga cgtgggtggag 240  
 gtgtgttcca aggactcgca ggtgtccggc gacggagggtt ggtggacccgg gcagctgaac 300  
 cagcgggtgg gcatcttccc cagcaactac gtgaccctgc gcagcgccctt ctccagccgc 360  
 tgccagcccg gccggcgagga agaaattgtt tttccggagc tcaccttggaa agagattttt 420  
 ggcacatgggg gctttggggaa ggtctatgtt gctttcttggaa tagggatgtt ggttgcgtg 480  
 aaacgcgtcc gcccacggaccat tcatggggatc ccatagagaa tgttccaa 540

gaggccaagc tcttcgccc gctgaagcac cccaacatca ttgcctaag aggggtatgt 600  
 ctgaaggagc ccaacctctg ctggcatg gagttgctc gtggaggacc tttgaataga 660  
 gtgttatctg ggaaaaggat tccccagac atccctgtga attggctgt gcagattgc 720  
 agagggatga actactact ttagtggca attgttccca tcatccaccg cgacctaag 780  
 tccagcaaca tattgtatcc cagaagggt gagaatggag acctgagcaa caagattctg 840  
 aagatcaactg atttggct ggctcggaa tggcaccgaa ccaccaagat gagtgcggca 900  
 gggacgtatg ctggatggc acccgaagtc atccggcct ccatgtttc caaaggcagt 960  
 gatgtgtggc gctatgggt gctactttgg gatgtgtggc gatgtgtggc gctatgggt 1020  
 ggcattgtatg gcttagcagt cgcttatggc gatgtgtggc gatgtgtggc gctatgggt 1080  
 ctcttacgt gcccagaacc tttgccaaa ctcatggaa actgtggaa tcctgatccc 1140  
 cactcacgac catcttcac gaatatcctg gaccagctaa ccaccaataga ggagtctgg 1200  
 ttctttgaaa tgcccaagga ctccctccac tgccgtcagg acaactggaa acacgagatt 1260  
 caggagatgt ttgaccaact caggccaaa gaaaaggaaac ttgcacccg ggaggaggag 1320  
 ctgacgggg ctgcaactgca gcagaagaac caggaggaac tgctcgccg tcgggagcag 1380  
 gagctggccg agcgggagat tgacatcctg gaacgggagc tcaacatcat catccaccag 1440  
 ctgtgcaggc agaagccccg ggtgaagaaa cgcaaggcga agttcaggaa gagccggctg 1500  
 aagctcaagg atggcaaccg catcagccct ccttctggg tccagcaca gttcaagggt 1560  
 caggccctccc ctaccatggc taaaaggaa agtcttatac acagccgtc cagtcctcc 1620  
 gcaagccccc ccatcattcc tcgccttc gccatccatg gtgagactgt tagcaaaacc 1680  
 tggggcagga gctcagtctg cccaaaggag gaaggggagg aggaggagaa gagggcccca 1740  
 aagaagaagg gacggacgtg gggccaggg acgcttggc agaaggagct tgctcggg 1800  
 gatgaaagcc tcaagtccct ggtagatggc tataaggact ggtcgccag tgcccccaac 1860  
 ctggtaagg gcccaaggag tagcccgcc ctgcccagggt tcaccagcct tatggagatg 1920  
 ggttaagttca cagaggatga ggacagtgtaa ggcccaggga gtggagagag tcgcctacag 1980  
 cattcaccca gccagttca cctctgtatc ccattccctc gtggagagga tggcgatggc 2040  
 ccctccaggta atggaatcca tgaggagccc accccagtc actcgccac gagtacccct 2100  
 cagctgacgc caaccaacag cctcaagcgg ggcgggtccc accaccggc ctgcgaggtg 2160  
 gctctgtcg gctgtggggc tggttggc gccacaggcc tagggttgc ctgtctggaa 2220  
 gctggcaagt gcccagtctg tcccctggg gacccgtc caccagcccg ggaggagaag 2280  
 aaaaagacggg agggtcttt tcagagggtcc agccgtcctc gtcggagcac cagccccc 2340  
 tcccggaaacg ttttcaagaa ggaggagccc atgctgttc taggagacc ctgtccctcc 2400  
 ctgacgctgc tccctctcc ctccatctcc gagtgcaact ccacacgctc cctgctgc 2460  
 tcggacagcg atgaaattgt cgtgtatgag atgcccgtc gcccagtctg gcccctccc 2520  
 ctgagttccat gtacccacaa cccctggc aatgtcccgat tagagcgtt caaacgagat 2580  
 cctaaccat ctctgactcc caccatgtc accctcacca cccctcgca gcccagcgt 2640  
 caccggcggg ctcccttc tggtggccctt cccagttccca gccgagaccc aggtgaattc 2700  
 ccccgctctcc ctgacccaa tggtgttcc cccccaaccc caaggcgtc gaacactcag 2760  
 caggactcta ctttggagag acccaagact ctggagttt tcgcctcgcc gcgtccttct 2820  
 gccaaccggc aacggcttgc cccttgggg ttgtgtcccc ccagccatgc ccgcagcacc 2880  
 tccccagccca acagctccag cacagagacg cccagcaacc tggactctg cttgtctagc 2940  
 agtagcagca ctgttagagga gccccttgc cttccagccc tgctccctt ccaggcagg 3000  
 ccgctgcccc cgactgagcg gacgtccctg gacctggatc cagagggca gagtcaggac 3060  
 agcaccgtgc cgctgtgcag agcggaaactg aacacacaca ggcctgcccc ttatgagatc 3120  
 caggcaggat tctggctta g 3141

<210> 28  
 <211> 1244  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474560CB1

<400> 28  
 gtcttattgc caagataatt acaactggag aattggcccc acaggaaaca acaattacag 60  
 agataaaaca aaaattgtat caaatacctg atgaagaggg cattgttattt gatggatttc 120  
 caagagatgt tgccctggc ctatcttttgg agggccatctt ctgtacccccc gattttgggg 180  
 tatttttttttggc ttgtgtctat cagagactca aaaaaggatc actgaagcgt gcagaacacg 240  
 agggccgacc agacgacaat gtaaaaagctt cccaaaggag actaatgaac ttcaaggcaga 300  
 atgctgtccat atgggttttttggc ttgtgtctat cagagggatc actaatgaac ttcaaggcaga 360  
 accgcgtatca ggatgtggatc ttctatgtatc tcaagcatggc agttgacaaac aagtttatttc 420  
 caaacaaga ggtctcaggatc ggttcaagtg accttgcatttcc ttgcgtatc 480

gagagatcat	tgatacagga	tctgattatg	aagatcaggg	tgtatgaccag	ttaaatgtat	540
ttggagagga	cactatggga	gttttcatgg	aagatttgag	aaagtgtaaa	attatttca	600
taatttggtgg	tcctggctct	ggcaaaggca	cacagtgtga	aaagctggtg	aaaaaatatg	660
gatttacaca	tctctcaact	ggcgagctcc	tgcgtgagga	actggcatca	gaatctgaaa	720
gaagcaaatt	gatcagagac	attatggaac	gtggagacct	ggtgcctca	ggcatcggtt	780
tggagctct	gaaggaggcc	atggtggcca	gcctcgggga	caccaggggc	ttccctgattg	840
acggctatcc	tcggaggtg	aagcaagggg	aagagttcgg	acgcaggatt	ggagaccac	900
agttgggtat	ctgtatggc	tgctcggcag	acaccatgac	caaccgcctt	ctccaaagga	960
gccccggacag	cctgcctgt	gacgcaccca	ccaagaccat	cgcgaagcgc	ctagaagcct	1020
actaccggac	gtccatcccc	gtgatgcct	actacgagac	aaaaacacag	ctacacaaga	1080
taaatgcaga	gggaacacca	gaggacgttt	ttcttcact	ctgcacagct	attgactcta	1140
ttattttctg	aaggaaaaaa	tgcgtttag	aatggaaaca	aaaaaacatt	aaaagttca	1200
ttccgttaac	acaatgtttc	aagttaaacc	ttttgtgtca	ccgc		1244

<210> 29  
<211> 1661  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7474602CB1

<400> 29

gcggccggcg	cgagagcgaa	agagggaaact	gcagaggagg	aagctgcgcc	gcagcccgag	60
ccgccccggc	tccccggcgc	ctctgcgc	gcgcggcgcc	cccgccgc	cctcccccagc	120
gcgccccccg	ccgctccctc	gcgcgcgcgt	cgtcgccat	ggcccgggag	aacggcgaga	180
gcagctcctc	ctggaaaaag	caagctgaag	acatcaagaa	gatcttcgag	ttcaaagaga	240
ccctcggaac	cggggcctt	tccgaagtgg	tttagctga	agagaaggca	actggcaagc	300
tctttgtctgt	gaagtgtatc	cctaagaagg	cgtcgaaggg	caaggaaagc	agcatagaga	360
atagatagtc	cgtcctgaga	aagattaagc	atgaaaatat	tgttgcctgt	gaagacattt	420
atgaaaagccc	aaatcacctg	tacttggtca	tgcagctggt	gtccgggtgg	gagctgtttg	480
accggatagt	ggagaagggg	ttttatatacg	agaaggatgc	cagcaactctg	atccgccaag	540
tcttggacgc	cgtgtactat	ctccacagaaaa	tgggcacatcg	ccacacagagac	ctcaagcccg	600
aaaatcttt	gtactacatg	caagatgggg	agtccaaaat	aatgtacatgt	gactttggat	660
tgtcaaaaat	ggagggcaaa	ggagatgtga	tgtccactgc	ctgtggaaact	ccaggctatg	720
tcgctcctga	agtctctcgcc	cagaaaacctt	acagcaaaagc	cgtgtactgc	tggtccatcg	780
gagtgattgc	ctacatcttgc	ctctgcggct	accctcctt	ttatgtgaa	aatgactcca	840
agctcttga	gcagatcctc	aaggcggaaat	atgagtttga	ctctccctac	tgggatgaca	900
tctccgactc	tgcaaaaagac	ttcattcgga	acctgtatgg	gaaggacccg	aataaaaagat	960
acacgtgtga	gcaggcgagct	cggcacccat	ggatcgctgg	tgacacagcc	ctcaacaaaa	1020
acatccacga	gtccgtcagc	gcccagatcc	ggaaaaactt	tgccaagagc	aaatggagac	1080
aagcatttaa	tgccacggcc	gtcgtgagac	atatgagaaa	actacaccc	ggcagcagcc	1140
tggacagttc	aaatgtcaagt	gtttcgagca	gcctcagttt	ggccagccaa	aaagactgtg	1200
cgtctggcac	cttccaaatgt	ctgttagttt	atttcttc	cgtcggggggt	ctcaggagg	1260
ggagccggag	ggagacccag	gcccacact	gtgacggcag	tgactctgg	aagcaagtga	1320
ctggccctgg	agggtggggcc	gggggtcggg	gctggggaaag	gggagccca	gggtcgccag	1380
agccgcgagc	cactcccgagc	agaccccccac	ttgcatgttgc	cccttcctg	catagactgt	1440
gaagaccgaa	gtttttttat	ggccatattt	tctactgcaa	ttcttgcgttgc	ttcatttctc	1500
acaaaactgt	ctgactcgag	gggcgtgtat	ttcataggat	ctgtgtgtgt	ataatacgat	1560
cttgc当地	tctactgaa	cgggacccat	tatttccttc	tccttaacacc	atcggttcca	1620
ctttctc当地	tgttagttaac	cgtctatgg	gtgtctttca	t		1661

<210> 30  
<211> 912  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7475509CB1

<400> 30

```

cgacgggtgg gctcggtccc ggcgctggc tgaggggagg gttgtctta aaagtcttc 60
cttccccctg taggggcggc cggcgagtc cagtgagac ggagggtgcc agaggtaggg 120
ggccgagaaa caaatttcc ggggttcct ccggggccgc gtcggggct gcgcgttga 180
ccgccccctt cctcgcgaag gcaatggct ccaaactctt ggcgcggc atcctcggtc 240
cgcccggtc gggcaagggc accgtgtgcc agaggatgcg ccagaacttt ggtctccagc 300
atctctccag cggcacttc ttgcgggaga acatcaaggc cagcaccgaa gttggtgaga 360
tggcaaaagca gatatagag aaaagtcttt tggtccaga ccatgtgtatc acacgcctaa 420
tgatgtccca gttggagaac aggctggcc agactggct cttgtatggt ttctcttagga 480
cattaggaca agccaaagcc ctggacaaaaa tctgtgtatggt ggatctatgt atcagtttga 540
atattccatt taaaacactt aaagatcgcc tcagccggc ttggattcac cctcttagcg 600
gaagggtata taacctggac ttcaatccac ctcatgtaca tggattgtat gacgtcaactg 660
gtgaaccgtt agtccagcag gaggatgata aacccgaagc agttgtgtcc aggctaagac 720
agtacaaga cgtggcaaaag ccagtcattt aattatacaa gagccgaggaa gtgtccacc 780
aattttcccg gaaccggaga cgaacaaaaa tctggccctt cgtttacaca acttttctca 840
acaagatcac acctattcag tccaaaggaaat cattttgacc tgcccatggaa gaccggaaatg 900
ggcctcactc ct 912

```

<210> 31  
<211> 2858  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7475491CB1

<400> 31

ggtacgagcc ggatcaacttgc tacggcgca ggtgtctggac agactctctc cttcccccctcc 60  
 ccttctacct cctctctggc cagctcagg tgcagttct cttgggaaact gtcacccctt 120  
 ccggagcagg ggaagctgcc cctgtccccgg gaggagcg ggcaccccg gccccccagga 180  
 cacgcgcgtgt gagtcccgcc gcggtgcgcc tgggaggaag ggggagggtcg gaggaggggg 240  
 caccgcggcc cccggtataa ggagcaaaagg acccggtctgc ccagtccctc atgatcatga 300  
 acaagatgaa gaactttaag cggcggttct ccctgtcagt gccccgcact gagaccattg 360  
 aagaatcctt ggctgaatttcc acggagcaat tcaaccagct ccacaaccgg cgaaatgaga 420  
 acttgcagct cggctcttcc ggcagagacc ccccgagga gtgcagcacc ttctcccaa 480  
 cagacagccg ggaggagccg gggcagctc cccctggctc gcagttccag cggcggcaga 540  
 accagcccg cttctccatg gaggtggggg ccttggcgc tctgccccgg 600  
 gatgcacca caaggggtgtg cacaggaggg cagtcgttcc acagccagac tttgacgtca 660  
 gcaagaggtct ctctctgccc atggatatcc gcctgccccca ggaatctca 720  
 agatggagag cccagatctg cccaagccgc tcaagcccat gtccccccgg 780  
 cagacatgg ctttggaaaa ctggaaacat acgtgaaact ggacaaaactg ggagagggca 840  
 cctatgccac agtcttcaaa gggcgcagca aactgacggg gaaccttgtg gcccctgaaag 900  
 agatccggct ggagcacgag gaggggagcgc cctgcactgc catccgagag gtgtctctgc 960  
 tgaagaacct gaagcacgcc aatattgtga ccctgcatga cctcatccac acagatcggt 1020  
 ccctcacccctt ggttttttag tacctggaca gtgacctgaa gcagtatctg gaccactgtg 1080  
 ggaacctcat gaggcatgc aacgtcaaga ttttcatgtt ccagctgctc cggggcctcg 1140  
 cctactgtca ccaccgcag atccctgcacc gggacctgaa gcccagaac ctgctcatca 1200  
 acgagaggggg ggagctgaag ctggccgact ttgactggc cagggccaaatc 1260  
 caaagactta ctccaatgag gtggtaccctc tgggtacag gcccccccgat gtgtctgtgg 1320  
 gatcccacaga gtactccacc cccattgata tggggcgt gggctgcata cactagcaga 1380  
 tggcccacagg gaggccccctc ttcccgggct ccacagtcaa ggaggagctg caccctcatct 1440  
 ttccgcctctt cgggacccccc acagaagaga cgtggcccg cgtgaccggc ttctctgag 1500  
 tccgcaccta cagttcccc tgcataccctc cgcagccgct catcaaccac gcccggcagg 1560  
 tggatacggg tggcatccac ctcttgagca gcctgtctct gtatgaatcc aagagtcgca 1620  
 tgcacggc ggtgtccctg agtcaacttct acttccggc tctgggagag cgtgtgcacc 1680  
 agcttggaaa cactgcctt atcttctccc tgaaggagat ccagctccag aaggaccagg 1740  
 gctaccgggg cttggccctt cagcagccag gacgaggggaa gaacaggccg cagagcatct 1800  
 tctgagccac gcccacccctt ctgtggccaa gggacaagag atcacatgg acaaaatc 1860  
 gggtaggatg gagccctgtgt gcccctcgga ggactgaaga acggggctg gcacaaatc 1920  
 tggaaagaccc cttggcagcc cttctggcca cggctgtttc ttctttgtgc acagccagcc 1980  
 cctcccccagt agccctcacc tgcataccaa cccctccctt acccacgttgg gggctggcat 2040  
 aagctgttcc cctgagagga catgaggggg gggcggtct cgttaccctt cccaccctgg 2100  
 tggggca cctgctgggg atgcacacgg atgacagaat caaggcgcca ggatggggcac 2160



```

ggccagcggag tataactgggg gtggcccaga gtccagcgcac aaggggcgcacc ctttcgtctac 2460
cctgtctgca cgtccccagca cccagccggag gccagactct tgggggtgagg acaactggga 2520
gggcctcgag actgacagtc gacagggtcaa ggctgagctg gccccggaaaga agcgcgagga 2580
gcggcggcgg gagaatgggagg ccaaaccgcgc cgagaggaag gtggccaagg gccccatgaa 2640
gttgggggcc cggaaagctgg actgaaaccgt ggcgggtggcc cttcccggtct gcggagagcc 2700
cgtccccacag atgtattttat tgtacaaaacc atgtgagcccc ggccggccca gcccaggccat 2760
ctcacgtgta cataatcaga gccacaataa attctatttc aaaaaaaaaa aaaaaaaaaa 2817

```

<210> 33  
<211> 5305  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7474496CB1

<400> 33  
atggctgggg gccgtggggc ccccgccgc ggccggacg agcctccgga gagctacccg 60  
caacgacagg accacgagct acaggccctg gaggccatct acggcggca cttccaaagac 120  
ctgccccgg acgcttgcgg acccgtcaaa gagccccctg aaatcaattt agttttgtac 180  
cctcaaggcc taactggtga agaagtataat gtaaaaagtgg atttgagggt taaaatgcccc 240  
cctactatc cagatgtagt tcctgaaata gaggtaaaaa atgccaaagg tctatcaattt 300  
gaaagtgtca atttggtaaa attcgcctt gaagaactgg ccaagaaaaca ctgtggggag 360  
gtgtatgatc ttgaactgg ttaccacgtg cagtcattt tcagcgagca taacaagccc 420  
ctcccaaggta cttttcatga agaaaatgtg gaaaggccgg ctcaagagga gcaacagagg 480  
ctgtggagg cccaaaggaa aagaagaaga gggcaggcac aacaacgtga aatcctgtcat 540  
gagattcaga gaaggaaaga agagataaaa gaagagaaaa aaaggaaaga aatggctaaag 600  
caggaacgtt tggaaattgc tagtttgc aaccagatc atacctctaa gaaggaccca 660  
ggaggacaca gaacggctgc cattctacat ggaggcttc ctgactttgtt agggaaatgg 720  
aaacatcggg caaactccctc aggaagggtct aggccgaaac gtcagtttgc ttgtatgtat 780  
agtgaagatt ctcctggctc ttgtgaaatt ctgtatttca atatggggag tcctgtatcat 840  
ctcattgtgc acaaaggaa atgtattggc agtgatgaac aacttggaaa attgtctac 900  
aatgttttg aaacagccac ttgtggctt gtcttgcgtt atgagtgggt cttcagtgg 960  
cagaaaaaaa tgggttccatt ctttaccagt aagaaaaaaag agaagattga taagtgc 1020  
aagcagattt aaggaacaga aacagaattc aactcaactgg taaaatttag ccatccaaat 1080  
gtgtacgct accttgcattt gaatctcaa gggcggacttggc acattgtgg 1140  
tttagtggagc acattgtgg ggtctcttt gctgcaccc tgagccactc aggccccatc 1200  
cctgtgcattt agtttgcagc gtacacagct cagctctgtt caggcccttga ttatctgcac 1260  
tgcatttgc tggtcataa ggtctgtggc gcatatcgat ttttgcgttgg 1320  
accgtcaaga ttacggacta tagcatttct aaggcccttgc cagacatttgc 1380  
ttgtttgagc aaaccccgatc tcgttttagt gacaatgttc tgccttataa aacggggaaag 1440  
aaaggatgtt tttttgcgtt tggccttctg ctgtgttcc ctagccaagg acaggatgt 1500  
gtggatgttcc ctgtgaccat ccctgttgcactt ttttgcgttgg 1560  
gtgtgtgtgt tggatgacaa gggaaatgttggc agtccccagc agttgttggaa acacagctt 1620  
taaatcccccc agccaaaaat gccttagtgc gggccgttgc acggatgttcc ttttgcgttgg 1680  
attatgttttgc agactgttgc tccatgttgc acggatgttcc ttttgcgttgg 1740  
cacagagac agttttcccg atacttcatt gggccgttgc acggatgttcc ttttgcgttgg 1800  
gagcttttgc gagctgtcat caaggtgcgg aacaaggatgg acggccgttgc ctacgcgtt 1860  
agcgcattcc ccatcaaccc ggccaggccgg cagttccgcga ggtcaagggg cgaagtgc 1920  
tgcgttgcac ggctgcacca tgagaacatt gtgcgttgc acggatgttcc ttttgcgttgg 1980  
acggccgc cggccggacc gggacgcggc cccccggactt ccggccgttgc acggatgttcc 2040  
accggatgttcc acggccggca gccggcgacg gacacagacg gcctggacag cgtatggcc 2100  
cccgccgc caccatctt cagcagctcg gtggagtggc gcacttcggg cgagcgttgc 2160  
ccagggtcc gttcccccgc caccggccgg ggctcccgatc atgacgagga cgacgacgg 2220  
acgagcactg gtggcgctt ctccatgttgc ttccatgttgc cttcagatc tgaaatgtat 2280  
ttatctttgc acaatgttgc tgagaaatgttgc aaaaatgttgc atcaggatgttgc 2340  
aaaagaatgttgc gctggccatgttgc aagtggatgttgc tgaggatgttgc acggatgttcc 2400  
acatccaga tggatgttgc tgagaatgttgc acgttgcgttgc acaccatgttgc 2460  
atcgagacaa cccgttgcactgttgc tgaggatgttgc ttccatgttgc atgacgagga 2520  
ccatgttgc aaggaaatgttgc tccatgttgc tgaaatgttgc acggatgttcc 2580  
tgaggatgttgc tgaaaatgttgc tgatgttgc acggatgttcc ttggatgttgc 2640  
cagcaacaaac aagacgttgc gacaggatgttgc tgatgttgc acggatgttcc 2700

<210> 34  
<211> 3269  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1834248CB1

<400> 34

```

ccgggacccg gagaagatgt cttcgcggac ggtgctggcc ccggggcaacg atcggaaactc 60
ggacacgcat ggcacacctgg gcagtggccg ctcttcggac aaaggcccgt cctggttccag 120
ccgctcactg ggtggccgtt gccggaactc catgcctcc tgttccgagg agcagcccca 180
cgtggggcaac taccgcctgc tgaggaccat tggaaaggc aaccttgcca aagtcaagct 240
ggctcggcac atccctcaactg gtcgggaggt tgccatcaag attatcgaca aaaccccgag 300
gaatccccagc agcctgcaga agctgttccg agaagtccgc atcatgaagg gccttaaaaaacca 360
cccccaacatc gtgaagctct ttgagggtat tgagacttag aagacgctgt acctggtgat 420
ggatgacgca agtgtctggag aagtgtttga ctacctcggt tcgcatggcc gcatgaagg 480

```

gaaggaaagct	cgagccaaagt	tccgcacagat	tggttcggct	gtgcactatt	gtcaccagaa	540
aaatattgt	cacaggggacc	tgaaggctga	gaacctcttg	ctggatgccc	aggccaaacat	600
caagattgt	gactttggct	tcagcaacga	gttcacgctg	ggatcgaagc	tggacacgtt	660
ctgcggggac	cccccatatg	ccggcccgga	gctgtttcag	gcaagaagaat	acgacggggc	720
ggagggtggac	atctggagcc	tgggagtcat	cctgtacacc	ctcgctcagcg	gtccctgtcc	780
cttcgcacgg	cacaacctca	aggagctgctg	ggagcgagta	ctcagagggg	agtaccgggt	840
ccctttctac	atgtcaacag	actgtgagag	catcctgcgg	agatttttgg	tgctgaaccc	900
agctaaacgc	tgtactctcg	agcaaatcat	gaaagacaaa	tgatcaaca	tcggctatga	960
gggtgaggag	ttgaagccat	acacagagcc	cgaggaggac	ttcggggaca	ccaagagaat	1020
tgaggtgatg	gtgggtatgg	tgtacacacg	ggaagaataa	aaagactctt	tgaccagcca	1080
gaagtacaac	gaagtgaccg	ccacacct	cctgctggc	aggaagactg	aggagggtgg	1140
ggaccggggc	gccccagggc	tggccctggc	acgggtcgcc	gcgcccacgc	acaccaccaa	1200
cggaacaagt	tccagcaaag	gcaccagcca	cagcaaaggg	cagcggagtt	ccttgcaccc	1260
ctaccacccgc	cagcgcaggc	atagcgattt	ctgtggcca	tcccctgcac	ccctgcaccc	1320
caaacgcagc	ccgacgagca	cgggggaggc	ggagctgaag	gaggagcggc	tgccaggccg	1380
gaaggcggagc	tgcagcaccg	cggggagttgg	gagtcgaggg	ctgccccctt	ccagccccc	1440
ggtcagcagc	gccccacaacc	ccaacaaggc	agagatccca	gagcggcgga	aggacagcac	1500
gagcacccccc	aacaacctcc	cctctagcat	gatgaccgc	agaaaacact	acgtttgcac	1560
agaacgcccc	ggggctgagc	ccccgtccct	gttgcacaaat	ggggaaagaaa	acagctcagg	1620
caccccacgg	gtggccctctg	cctcccccct	cagtcacagc	ctggcacc	catcagggga	1680
gcgaggccgc	ctggcacgcg	gttccacat	ccgcacacc	ttccatggtg	gcggagtcgg	1740
ggaccggccgg	gcaggggggt	gggggtgtgg	gggtgtcgag	aatgggcccc	ctggctctcc	1800
cacactggcc	catgaggctg	cacccctgccc	ccccggccgg	ccccggccca	ccaccaacact	1860
cttcaccaag	ctgacacctca	aactgacc	aagggtcgca	gacgaaac	agagaatcgg	1920
gggacctggag	gtcacaagg	gccatctacc	ttgggatcaa	acgaaacccg	ccccccggct	1980
gctccgat	ccctggagtg	tgaagctgac	cagctcgcc	cctcttgagg	ccctgtatg	2040
agctctgcgc	caggccacac	cagccggcccg	ctggccgtc	cgccagccac	agccgttct	2100
gtgtgcgtc	ctgcacgggg	gtggccggcg	gcccggcc	ctgtcccact	tcgaagtgg	2160
gggtctgcac	ctgccccggc	caggcttgcc	gggagttctc	ttccggccgt	tggccggcac	2220
cgcctggcc	ttccgcaccc	tgcgcaccc	catctccaa	gac	tctgagccac	2280
cacggtccca	gggccttac	tcttcctctc	ccttgc	ttacttctaa	caggagg	2340
aggggc	cagg	ggggattc	ttcccttatac	atcacctc	ttatattttgg	2400
gggcaaaat	tg	ttccctct	gtgttctct	ggggccg	ttatattttgg	2400
gggtctgc	ggggagctg	gcacccct	ggggccg	agcacagaag	aaaggatgg	2460
ccctaccaag	agggcac	cttcc	ggggcc	ggccatcttgc	ttccctct	2520
gaggaaatct	tccatctc	ccaaacgctc	aaaattag	cttggg	ggcagg	2580
gctgtgt	ctaaagact	gagaatctg	gggactgg	gtgggg	gaga	2640
ttccctcccc	tccgtcccc	tcacgtcaa	acccca	cctgc	agg	2700
ggcactttgt	acaatcc	gtaaatacc	cacccctcc	cctctgc	gg	2760
ggagctggc	ctgtcac	cggtttttaa	gttataac	ccgac	ct	2820
ccccctcacc	tgca	ccataa	atttagg	gtcccc	ct	2880
gacccttag	tttcc	ctgc	atttagg	gtt	ct	2940
atccaggc	tggtt	ttc	ggat	aaagaag	ggc	3000
tggggttgg	aaggta	tgtcgt	agaaggc	atggaaact	ggc	3060
cgaggcccc	gttcc	tat	aggc	atggaaact	cc	3120
ttacactg	ctcta	ccgt	ctgt	tttattca	aaacaat	3180
tgacccat	aaaa	aaaa	aaaa	aaaa	aaaa	3240
						3269

<210> 35  
<211> 3017  
<212> DNA  
<213> *Homo sapiens*

<220>  
<221> misc\_feature  
<223> Incyte ID No: 71584520CB1

<400> 35

```

gggcaagccg  ctggcgatgc  tggaggttcg  ctagccgaag  cggctgcata  tggcgcccg  60
tctggcccccgc  gtgctcgagg  cggattctgc  ccggccgtccc  cggagccctc  ggcggcccccgc  120
tgagccccgc  gatcaacttcc  tccctgtac  caaccggcgc  tgccaggtag  agcctggcaa  180
tgcgcgtttgg  gtgtgtgact  ctgggcgaca  agaagaacta  taaccagcca  tggggaggta  240
ctgacagata  tgatttggga  cagggtcatca  agactgagga  gttttgtgaa  atcttccggg  300

```

ccaaggacaa gacgacaggc aagctgcaca cctgcaagaa gttccagaag cgggacggcc 360  
 gcaagggtcg gaaagctgcc aagaacgaga taggcattct caagatggg aagcatccca 420  
 acatcctaca gctgggtggat gtgtttgtga cccgcaagga gtactttatc ttccctggagc 480  
 tggccacggg gaggggagggtt gttactggta tcctggacca gggctactac tcggagcgag 540  
 acacaagcaa cgtggtacgg caagtctgg aggccgtggc ctatggcac tcactcaaga 600  
 tcgtgcacag gaatctcaag ctggagaacc tggttacta caaccggctg aagaactcg 660  
 agattgtcat cagtgacttc catctggcta agtagaaaa tggctctatc aaggagccct 720  
 gtgggaccctc cgagtatctg gccccagagg tggttagggcg gcagcggtat gacgcctg 780  
 tggactgtcg ggcattggta gtcatacatgt acatcctgtt ttcaggcaac ccacccctt 840  
 atgaggagggtt ggaagaagat gattatgaga accatgataa gaatctctc cgcaagatcc 900  
 tggctgggtga ctatgagggtt gactctccat attgggtgta tatttcgcag gcagccaaag 960  
 acctggtcac aaggctgtatg gagggtggagc aagaccagcg gatcaactgca gaagaggcca 1020  
 tctcccatgaa gtggattctt ggaatctgtt ctctgtataa gaacatcaag gatgggtgtct 1080  
 gtgcccagat tgaaaagaac tttggccaggg ccaagtggaa gaaggctgtc cgagtggacca 1140  
 ccctcatgaa acggctccgg gcaccaggagc agtcccgacac ggctgcagcc cagtcggcc 1200  
 cagccacaga cactggcacc cccggggctg cagggtggc cacagctgca gtcgcgagtg 1260  
 gagctacccctc agccctgtgg ggtgtatgtc tctgtgtctc aaagagtgta aatgtggccc 1320  
 cccgagaccc tagtggccacc ccagccacag atggaaatgtc accccccggc actgtatggca 1380  
 gtgtcaccctc agccaccgtt ggaagcatca ctccagccac tgatgggaggt gtcacccctc 1440  
 ccactgacacag gagegtactt ccagccactg atggggagagc cacaccaggc acagaagaga 1500  
 gcaactgtgccc caccacccaa agcaatgtccca tgctggccac caaggcagct gccacccctg 1560  
 agccggctat ggcggccggc gacacgaccc ccccaaggagg cgccacaggg caggctccac 1620  
 cctctagtaa aggggaagag gctgtgtttt atgcccaggta gtctcaaagg gaggaggcca 1680  
 gctgagtagg cagcctggc agggggggca gggggatgggc aggagggtgg gagagtggat 1740  
 gaggggcttc tcactgtaca tagaggtact ggcataatgtc cctcgatccc ccatgcccccc 1800  
 acatcccagt ggggcataac taggggtcac ggggatgggc tctcgatccc tgggtgtatg 1860  
 tgggtgtatg gttggcaggc cagtgccagg gccggcccca gcccctgcat ggattccctg 1920  
 tggcttttctt gtcttttgc agtttcacca gtttctgtc ctgtggat gctgtcttag 1980  
 ggataactcg ggggctctg ctctccttcc ctttcccttc ttgcctcacc attcccttag 2040  
 gcaggccctg caggtccac actctccctg gcccctaaact tggggccct tgccctgaga 2100  
 gctggcttc cagcgaggcc ctgtcagccggtt ctctaggctc ctgcacatga aggtgtgtc 2160  
 ctgtgggtgtt tgggtgtctc taggagcaga tacaggctgg tatagaggat gcagaaagg 2220  
 agggcagttt gtttaagtcc agacttggca catggctagg gataactgtc actagctgt 2280  
 gaggtcctca ggagtggaga gaaatgttagt gaggggcagaa gcttccattt ttgccttcc 2340  
 taagaccctgtt tattttgtgtt tatttcctgc ctttccggat cctgcactgg gctgcccctgt 2400  
 accctgaacc tcatgagcc ttaaggaaaa ggaggaacaa ttaggacgtg gcaatgagac 2460  
 ctggcaggcc agagtacaag cccagccccc agtgcctccag ccttacttgg tccttaccct 2520  
 gggccaaacaa gggggctg atacccctt gcttccctt gatgcccacc tcctacaatc 2580  
 tcagccacca agtcccttc acccttagggg gcttgcgtca tggcaataac tcataatctg 2640  
 atttgggtt tggcccttta cagggcaga ttttctgtc agttcaacaa tgaatgaag 2700  
 aggaactccc tctttctaca gtcacttctc atcagaggcc caggtgcctc agagccacat 2760  
 ttagttgtt tttctggat gaggaaatgtt gtttaactccc cccagttcc tgagggagggc 2820  
 tcctgacagg tggcccttgc tggcccttgc tggcccttgc tggcccttgc tggcccttgc 2880  
 tcggcccttgc tcggcactcc gtgggtgtcc tggcccttgc tggcccttgc tggcccttgc 2940  
 ctctgggtgtt tgaagggtcg tgggttaact gtgtgcctac tgaacctggc aaataaaacat 3000  
 caccctgcaaa agccaaaa 3001  
 3017

<210> 36  
 <211> 2168  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7475538CB1

<400> 36  
 gctgtcatcg ttccgtggc cctgcttgc ggcacgctct cggcgcatgc gttttttatg 60  
 cgggatataag cttgctgtc cgtgacagcg gagggtctagg aaaaggcgca gtggggcccg 120  
 gagctgtcac ccctgactcg acgcacgttc ctttcttctg gtgcgtggc ctacagggac 180  
 cgcggccatgtg tgcacgttgc ggcgtgtgc taggcaacag cgtgcgagct cagatcagcg 240  
 tgggggtggag gagaatgtt gtttggatg tcaaggccac agggggcacag gcccacgact 300  
 gcagcgggat ggaccaggatc tgcacccctgg gcccacatgg ggagggcgcc cacggaaatcg 360

tcttcaggc caagcacgtg gagactggcg agatagttgc cctcaagaag gtggccctaa 420  
ggcggttgc agacgcgttc cctaaccagg ccctgcggga gattaaggct ctgcaggaga 480  
tggaggacaa tcagtatgtg gtacaactga aggctgttt cccacacggg ggaggcttt 540  
tgctggccct tgagttcatg ctgtcgatc tggccgaggt ggtgcgccc gcccagaggc 600  
caactagccaa ggcacaggc aagagctacc tgcaagatgt gctcaagggt gtcgccttct 660  
gcccattgc caacattgtt catcgggacc tggatcgatc caactgttc atcagcgcct 720  
caggccagct caagatagcg gactttggcc tggatcgatc ctttccccca gacggcagcc 780  
gcctctacac acaccagggt gccaccagggt ggtaccgagc ccccgagctc ctgtatggtg 840  
cccggccaggta tgaccaggc gtcgatctgt ggtctgtggg ctgcattatg ggggagctgt 900  
tgaatgggtc ccccccttcc cccggcaaga acgatattga acagctttgc tatgtgttcc 960  
gcatacttggg caccggaaac cctcaagtct ggccggagct cactgagctt ccggactaca 1020  
acaagatctc cttaaggag caggtgcctt tggcccttgc gggaggctgt cctgacgtct 1080  
ctccccaggc attggatctg ctgggtcaat tccttctctt ccctcttcac cagcgcattcg 1140  
cagttccaa ggctcttcctc catcaatct tcttcacagc tccctgcctt gcccattccat 1200  
ctgagctgcc gattccctcag cgctctagggg gacctgcccc caaggcccat ccaggcccc 1260  
ccccacatcca tgacttccac gtggaccggc ctcttgagga gtcgtgttg aacccagagc 1320  
tgattcggcc cttcatctg gagggtgag aagttggccc tggtccgtc tgcctgttcc 1380  
tcaggaccac tcagtcacc tggatcttgc ccacctgcct ggcttcaccc tccaaggcct 1440  
ccccatggcc acagtggcc cacccacac cctggccctt agcccttgcg agggttggtc 1500  
tcgaggcaga ggtcatgttc ccagccaaga gtatgagaac atccatgtcg gcaaggagaga 1560  
ttcatggcct gtgctcggtg agccttacct tctgtgtgt actgacgtac ccatcaggac 1620  
agttagtctc gtcgcctgac aaggcctgca tatgcagaat gacgtgcct gccttggtgc 1680  
tgcttccccc agtgcgcctt cctggtcaag gagaagtgcg gagaaggatgg tgcctttagt 1740  
ttggaaactc aagtggagg aagatttggt ttggtttat tctcagagcc attaaacact 1800  
agttcagtat gtgagatata gattctaaaa acctcagggt gctctgcctt atgtctgttc 1860  
ctcccttcatt tctctcaagg gaaatggcta aggtggcatt gtctcatggc tctcggtttt 1920  
ggggtcatgg ggagggttagc accaggcata gccactttt ccctgaggga ctccgtgtg 1980  
cttcacatca ctgagcactc atttagaagt gaggagacca gaagtctagg cccaggatg 2040  
gctccagttg gggatccagc aggagaccct ctgcacatga ggctggttt ccaacatcta 2100  
ctccctcagg atgagcgtga gccagaagca gctgtgttatt taagggaaaca agcgttccctg 2160  
gaattaaat

2168